Journal of Ambient Intelligence and Humanized Computing (2023) 14:3441–3474
https://doi.org/10.1007/s12652-021-03479-0

ORIGINAL RESEARCH

A comprehensive survey on machine translation for English, Hindi and Sanskrit languages

Sitender[1,2] · Seema Bawa[1] · Munish Kumar[3] · Sangeeta[2]

Received: 24 October 2020 / Accepted: 1 September 2021 / Published online: 13 September 2021

Abstract
Transforming text from one language to another by using computer systems automatically or with little human interventions is known as Machine Translation System (MTS). Divergence among natural languages in a multilingual environment makes Machine Translation (MT) a difficult and challenging task. The purpose of this paper is to present a comprehensive survey of MTS in general and for English, Hindi and Sanskrit languages in particular. The state-of-the-art MT approach is Neural Machine Translation (NMT) which has been used by Google, Amazon, Facebook and Microsoft but it requires large corpus as well as high computing systems. The availability of MT language modeling tools, parsers data repositories and evaluation metrics has been tabulated in this article. The classification of MTS, evaluation methods and platforms has been done based on a well-defined set of criteria. The new research avenues have been explored in this survey article which will help in developing good quality MTS. Although several surveys have been done on MTS but none of them have followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach including tools and evaluation methods as done in this survey specifically for English, Hindi and Sanskrit languages.

Keywords Artificial intelligence · BLEU · Knowledge representation · Machine translation · NIST · Natural language processing · Systematic survey · Statistical machine translation

1 Introduction

Natural languages have shown a vital role in shaping human social behavior as they prepare the necessary mechanism for day to day communication among human beings (Fromkin et al. 2011). Natural Language Processing (NLP) comprises of three basic components: processing, understanding and generation (Allen 1995). NLP is a sub-domain of Artificial Intelligence (AI) and Machine Translation (MT) is one of the application of NLP. Machine Translation (MT) is a mechanism of translating the sentences of one language designated as Source Language (SL) into other language designated as Target Language (TL) with the help of computers (Hutchins 1995; Hutchins and Somers 1992; Slocum 1985). The translation may occur one-to-one, i.e. from one SL to another TL, known as bi-lingual translation; one-to-many, i.e. from one SL into many TLs and many-to-many translation, i.e. from many SLs to many TLs known as Multilingual Machine Translation (MMT). MT comes under Natural Language Processing (NLP) domain which is a sub-domain of Artificial Intelligence (AI) (Rao 1998). The translation may be unidirectional or bidirectional. Several efforts have been made to review the MT systems whereas major contributions has been done by Antony (2013), Desai and Dabhi (2021), Garje and Kharate (2013), Naskar and Bandyopadhyay (2005). The research in the MT field has been increased rapidly in the last few decades. Therefore a systematic yet

✉ Sitender
sitender@thapar.edu; sitender@msit.in

Seema Bawa
seema@thapar.edu

Munish Kumar
munishcse@gmail.com

Sangeeta
sangeeta.phogat@gmail.com; sangeeta@msit.in

1 Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology, Patiala, Punjab, India

2 Department of Information Technology, Maharaja Surajmal Institute of Technology, New Delhi 110058, India

3 Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Batinda, Punjab 151001, India

critical evaluation of available MT techniques, methods and systems is needed. In this article, the authors have surveyed the traditional as well as state-of-the-art techniques and systems of MT. An effort has been made to identify existing MT approaches, development tools, data repositories, environments, evaluation metrics and platforms.

1.1 Motivation

According to Ethnologue languages of world, approximately 7102 languages and thousands of dialects have been used by people in the world (Lewis et al. 2015). Human translation has never been an effective solution for such problems due to less availability of human translators, high cost of manual translation and difficult to approach by everyone. According to Census of India 2001 data, 22 scheduled and 100 non-scheduled languages with approximately 1600 local dialects were being used by people (Dorr et al. 2004; Mallikarjun 2010). So, for the development of country like India, people have to exchange technology, science, ideas and work together without any language barrier. MT techniques can remove such problems in an effective manner. Thus, there is a great need of MT at the global level as well as local level in India also.

The summary of contribution and novelty of this review article is of many folds which are listed as follows:

- Presenting comparison of MT techniques and evaluation methods based on well-defined criteria to analyze the existing MT platforms with their characteristics and applications.
- Analyzed the availability of various language resources and presents word embedding techniques used in neural machine translation for Indian languages.
- Explored the new research areas in the field of machine translation for Indian languages.

1.2 Approaches of MTS

Figures 1 and 2 shows different MTS approaches (Dorr et al. 2004; Seasly 2003). Broadly we can categorize approaches into five groups: Direct Machine Translation (DMT), Rule-Based MT (RBMT), Corpus-Based MT (CBMT), Knowledge-Based MT (KBMT) and Hybrid Based MT (HBMT). RBMT is further divided into Transfer Based MT (TBMT) and Interlingua Based MT (IBMT) whereas CBMT is divided into Statistical MT (SMT) and Example-Based MT (EBMT). Neural Machine Translation (NMT) is an extension of SMT as depicted in Fig. 1. Figure 2 shows the level of complexity in different approaches in the form of Vauquois triangle. From bottom to top complexity increases.

1.2.1 DMT

DMT comes at the bottom of the triangle and needs fewer efforts. There is no intermediary representation of the source and target language, only word to word matching is performed for the translation and the system may have pre-processing and post-processing paring phases for the input sentence morphological analysis and the target sentence reordering, respectively. The system uses a bilingual dictionary for matching the SL words with TL words. Figure 3 depicts the DMT approach.

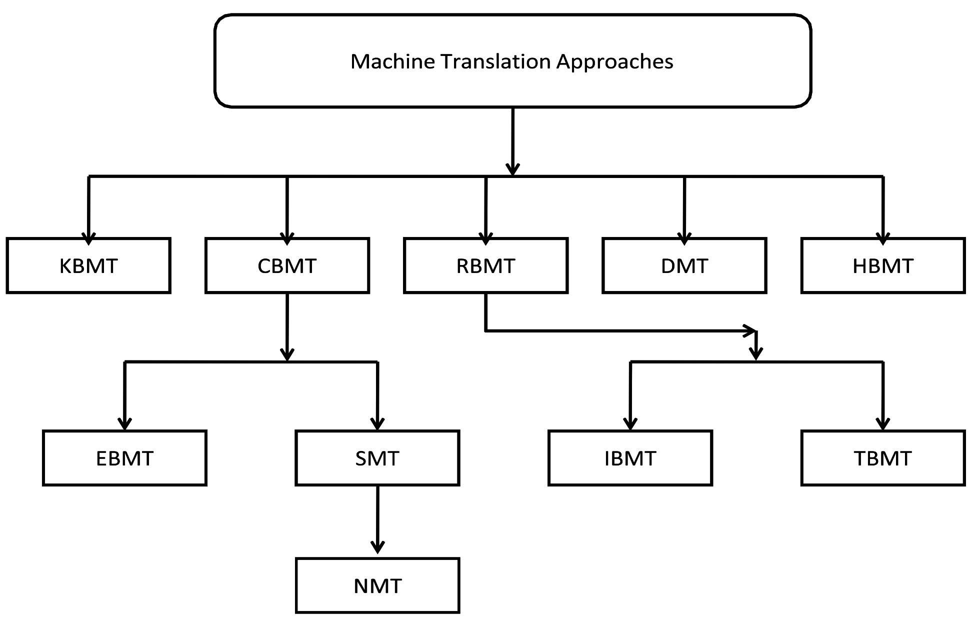

Fig. 1 MT approaches

1.2.2 TBMT

In this approach after the morphological analysis of input sentence, the syntactic and semantic analysis using the SL dictionary is performed to find out grammar structure and generates a parse tree. The system uses a set of transfer rules to transfer SL parse tree into TL with the help of a bilingual source-target language dictionary. The TL text is generated as per the grammar of TL using syntactic and semantic generator modules and the target language dictionary. The working of TBMT approach is depicted in Fig. 4.

1.2.3 IBMT

In this approach, SL text is analysed and an intermediate language independent code is generated to obtain the TL text. As the intermediate code representation is independent of SL as well as TL so could be used in multilingual machine translation. The language analyser is dependent on SL in the input process and the target language generator is dependent on the particular target language. The functioning of IBMT is shown in Fig. 5.

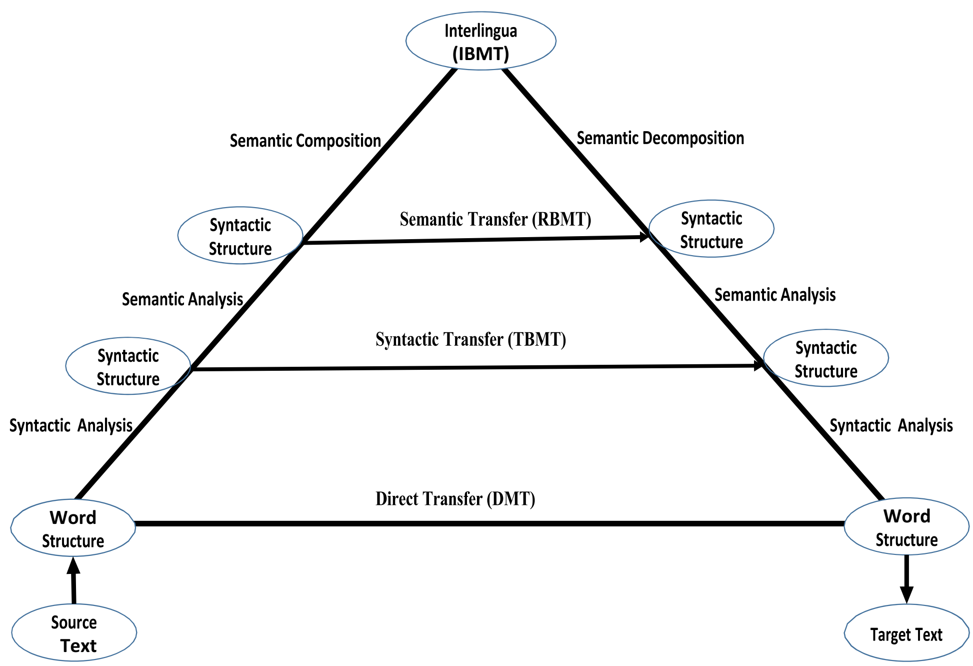

Fig. 2 Vauqois triangle

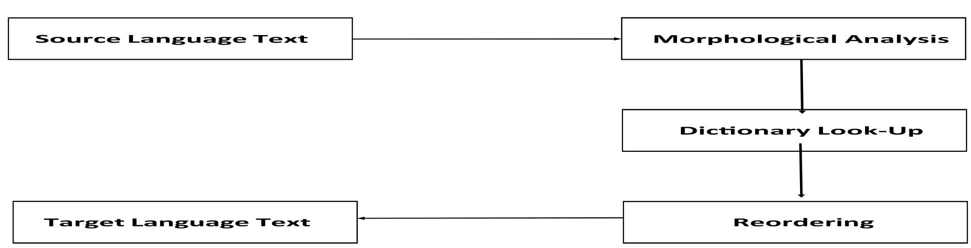

Fig. 3 Direct MT approach

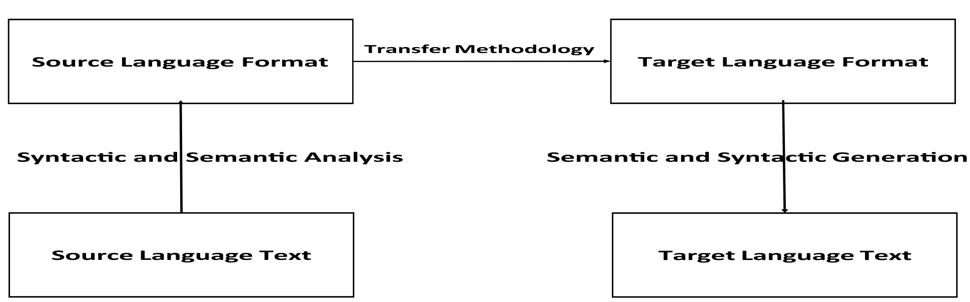

Fig. 4 Transfer based MT approach

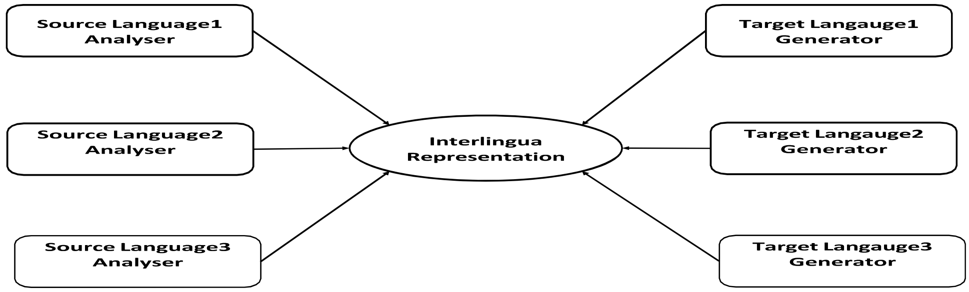

Fig. 5 Interlingua based MT approach

1.2.4 SMT

In this approach, statistical or probabilistic techniques have been applied in machine translation system development. There are two major components of this approach as-language model and the translation model. The language model produces the probability of occurrence for the strings of words in the source as well as the target language and also the conditional probabilities of occurrence of a word in the target language which translates a word in the source language. The multiplication of the probability of occurrence of a word in SL with the conditional probability of occurrence of a word corresponding to this word in TL provides the occurrence of source and destination pairs of words occurring in the corpus available for translation. This method requires a large amount of database and very complex statistical techniques to do the translation. The efficiency of the system increases with more training data sets and parallel corpora availability for the language pair. Machine translation can be done based on word, phrase, sentence, or hierarchical phrase. The translation model generally uses the N-gram model. N-gram model predicts the occurrence of the next word of the text given the previous words. The working process of the SMT approach is presented in Fig. 6.

1.2.5 EBMT

The basic translation principle used by this approach was analogy. This approach does not require huge amount of corpora, it needs a bilingual corpus of stored examples and using one of the matching algorithm to find the translation which matches with the source language sentence. Generally EBMT does not require any grammar rule base in detail; it uses only the stored examples and the matching algorithm to find the closest match corresponding to the given input sentence. The architecture of EBMT approach is shown in Fig. 7.

1.2.6 KBMT

This approach extracts the linguistic information from SL and stores that information into the knowledge base used for translation purpose. Information extraction is done by using bilingual dictionaries, language structure, stored translation information, domain specific information dictionaries etc. Figure 8 depicts the architecture of KBMT approach.

Each approach has its own advantages and disadvantages, so hybridization of two or more than two approaches might give a better translation quality. Hence researchers are focusing on hybridization of approaches at different levels for developing MTS. Comparison of MTS approaches have been done based on a set of well defined criteria as shown

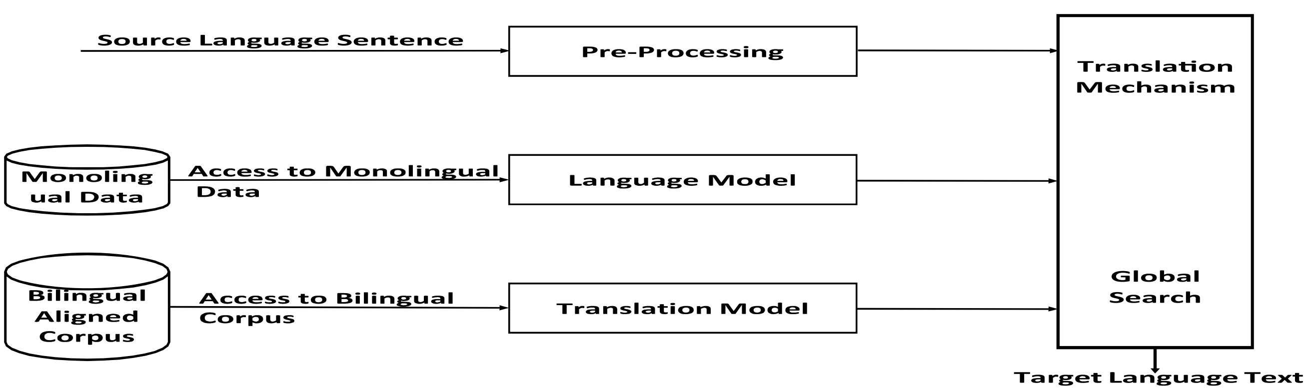

Fig. 6 Statistical MT approach

in Table 1. RBMT approach gives better results than other approaches, but needs deep linguistic knowledge, more time to create translation rules.

Corpus Based Machine Translation (CBMT) approach performs better than DMT for long sentence translation, but requires large volume of text corpus for both SL and TL, statistical tools, algorithms to handle and high computation power for the development of MTS. DMT approach is better for translating single clause sentences and requires less time to develop MTS. Neural Machine Translation is an emerging technique and reports similar results to the present state-of-art MTS (Hassan et al. 2018; Wu et al. 2016).

Hybridization of CBMT and RBMT can be done based on confidence-estimation and classification (Christopher and

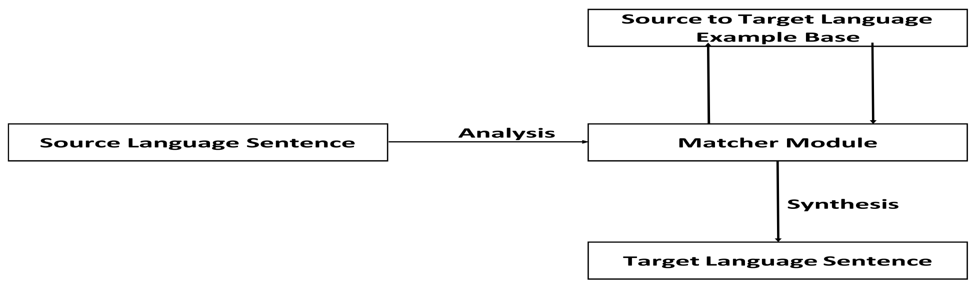

Fig. 7 Example based MT approach

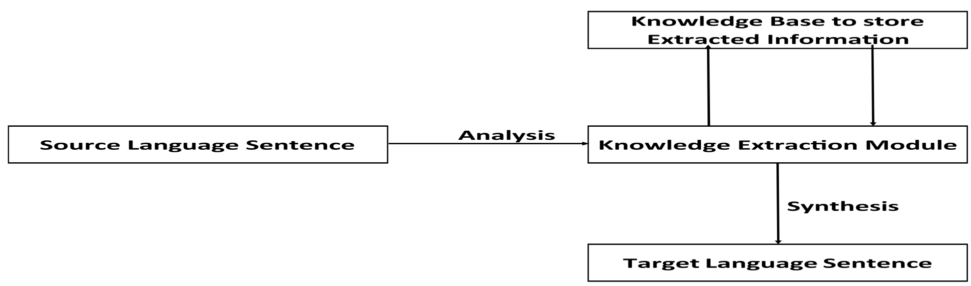

Fig. 8 Knowledge based MT approach

Table 1 Comparison of MT approaches based on several criteria

MT approach criteria	DMT	RBMT	CBMT	KBMT	NMT
Morphological analysis	Required	Required	Required	Required	Done by encoder
Syntactic and semantic analysis	Not required	Required	Required	Syntactic required not semantic	Encoder performs this task
Deep linguistic knowledge	Not required	Required	Not required	No, require inference engine	Training of encoder and decoder is required not simple, but less space is required than SMT
Simple to implement	Yes	No	Simple than RBMT	No	
Cost	Less costly	Costly in terms of time	Costly in terms of resources	Costly in terms of conceptualization	Costly in terms of computational power required (needs GPU)
Fast development	Yes	Time consuming	Faster then RBMT	Less than RBMT but less then CBMT	Once trained gives output in fractions of seconds
Efficiency	Better for simple and small translation	Most efficient	Better than DMT	Better than DMT and CBMT	Better than SMT
Large computation required	No	No	Yes	Yes	Yes
Word level translation	Yes	Yes	Yes	Yes	No
Sentence level translation	No	No	Yes	No	End-to-end translation

Rao 2010). However, the problem with such hybridization is the requirement of a large corpus of parallel sentences to extract translation rules to cover all aspects of natural language. To overcome such problems Recursive Chain-Learning (RCL) or Genetic Algorithms or Neural Networks can be used over the existing systems (Echizen-Ya et al. 2004). For translating fixed patterns, the RBMT approach was not effective, because conventional syntactic analyzers are not able to recognize such fixed patterns (collocation, idioms and compound nouns). To remove such problems specific pattern recognition modules can be added to the existing RBMT based systems. This will reduce the load on POS tagger and parser, helps in resolving word sense ambiguities (Jung et al. 1999). Other hybrid combinations are explained in Sects. 4.1 and 4.2.

The rest of the article is organized as Sect. 1 gives the introduction to MT, Motivation, the contribution of this article and approaches of MT. Section 2 describes the evolution of MT in general as well as for English, Hindi and Sanskrit languages. Section 3 explains the survey methodology adopted for the current work. Section 4 describes outcomes as results obtained from various MT systems. State-of-the-art MTS platforms, parsing and language modeling tools, available corpora have been discussed in Sect. 5. Section 6 highlights the role of Neural Networks in Machine Translation with some latest examples of MT systems based on NMT approach and Sect. 7 depicts MT evaluation methods and platforms with their characteristics. Section 8 provides research avenues generated from this work and recommendation for new researchers. Finally the concluding notes are given in Sect. 9.

2 Evolution of MTS

2.1 Evolution of MTS in general

Machine translation history had started in the 17th century when Discartes and Leibniz proposed the concept of mechanical dictionaries based on the method of universal numerical codes. But the actual proposal for the machine translation came in the 20th century. Figure 9 shows the development of machine translation in five phases in general (Hutchins 1995; Hutchins and Somers 1992).

2.2 MTS development in Indian perspective

The MTS development for Indian languages has started in 1990s and Fig. 10 shows various MTS developed for English, Hindi and Sanskrit languages based on different approaches.

The domain, efficiency, features and the research group associated with these MTS is explained in Sect. 4. Initially due to non-availability of online corpus for Indian languages compared to other languages, DMT and RBMT approaches have been used for developing MTS among Indian languages, although some CBMT based MTS for English to Indian languages or Indian to English language translation have also been developed. In 2003 the hybridization of different approaches have started for developing MTS. From 2009 to 2014 RBMT approach has been used extensively for MTS development. In the duration from 2016 to now the graph of CBMT increases due to the application of NMT approach in MTS. The hybrid approach was also used in parallel to RBMT and CBMT in a few MT systems during the same time. In hybridization, Artificial Neural Network (ANN) and Quantum Neural Network (QNN) techniques outperform compare to other combinations. RBMT approach dominates other approaches in Indian MT development scenario.

3 Survey process

The approach used for survey in this article follows the guidelines given in Budgen and Brereton (2006), Kitchenham et al. (2009), Moher et al. (2015). The different stages involved in the survey process are planning, execution, analysis of results, documentation of results and highlighting the research gaps. The planning of survey includes the creation of an effective research question framework as shown in Table 2, sources of articles as discussed in Sect. 3.1. Execution of survey includes criteria for searching the article as shown in Table 3, inclusion or exclusion criteria of articles in the survey.

3.1 Information sources

A broad perspective is essential for broad coverage of literature as suggested by Kitchenham et al. (2009) and Budgen and Brereton (2006). So the following electronic sources were used for searching the relevant articles for the survey:

- "Google Scholar (https://scholar.google.co.in/)"
- "IEEE Explorer (ieeexplore.ieee.org/)"
- "ACM Digital Library (dl.acm.org/)"
- "Science Direct (https://www.sciencedirect.com/)"
- "Springer (www.springerlink.com)"
- "ACL(https://www.aclweb.org/)"

3.2 Searching criteria

All the articles searched over electronic sources include the token" Machine Translation" which makes the process of searching relevant articles a time-consuming and challenging, as these articles are vast in numbers. So, a search

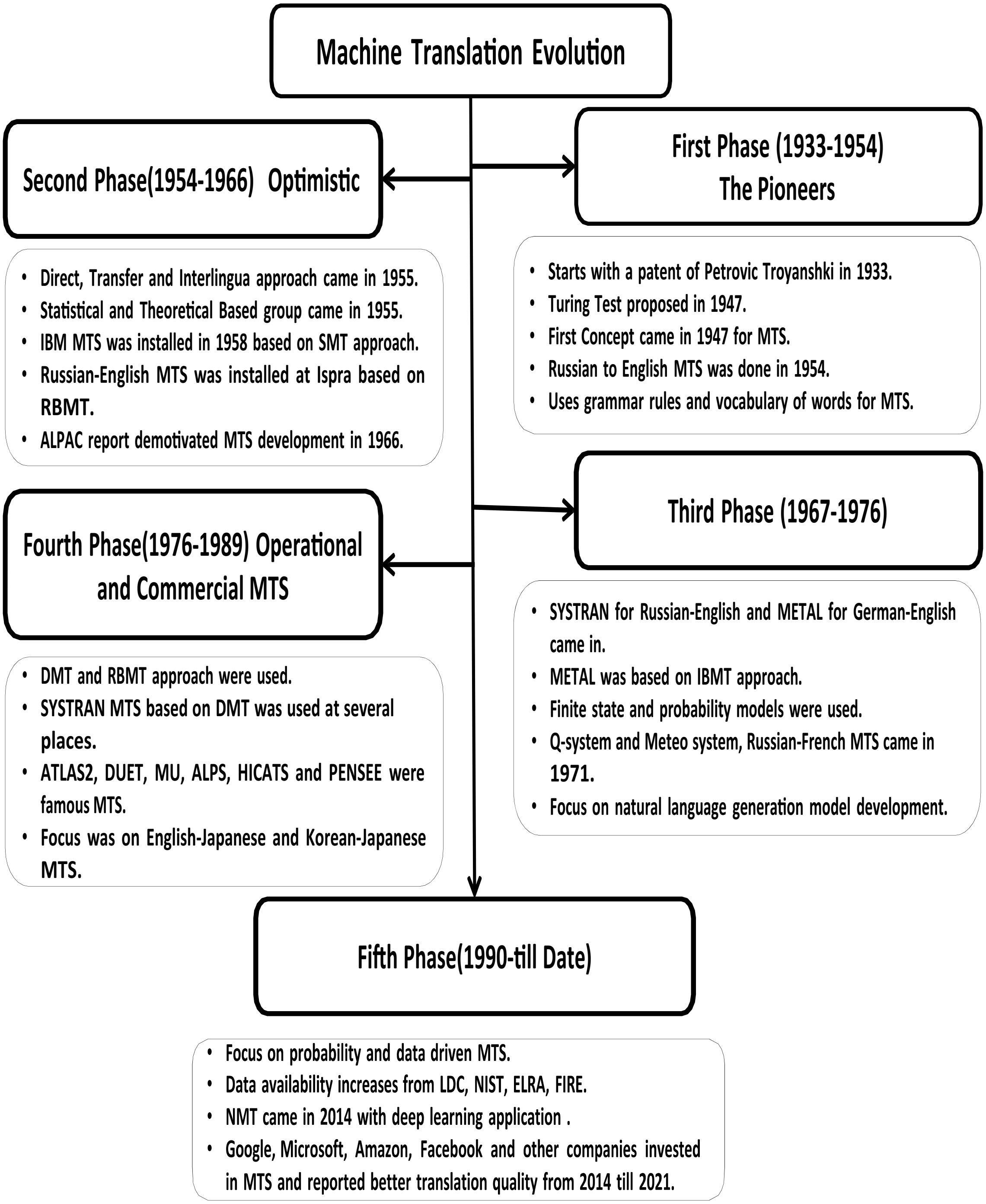

Fig. 9 MT evolution in general (Cho et al. 2014; Hutchins 1995; Hutchins and Somers 1992; Kalchbrenner and Blunsom 2013; Sutskever et al. 2014)

strategy is needed to include as many related articles as possible with ease and in less time. One such approach is presented in Table 3, but still, some of the right papers might not be added to this survey, a reason may be due to missing such keywords into the abstract part. The work on MT for Indian languages started in the 90s, and the current survey includes articles from different sources like journals, conferences, workshops, seminars, technical reports, and symposiums from 1990 to Feb 2021.

3.3 Inclusion/exclusion criteria

The process of including or excluding the article in the current survey is shown in Fig. 11. In the first phase, the

Fig. 10 Evolution of MT in Indian perspective based on different approaches

exclusion of articles has been done based on the title of the article. The exclusion percentage in this stage was 28%. In Phase-2, 1057 articles are separated from the original 1500 article database, and after studying their abstracts, only 410 articles are selected for the next phase based on their relevance to the field of machine translation. In Phase-3, after reviewing the full text of 410 articles only 220 are moved to the next phase, and rest are excluded. In Phase-4, the exclusion is done based on the MT for English, Hindi and Sanskrit languages and finally, 118 articles are included for the current survey.

Table 2 Research question framework

Sr. No.	Research questions	Motivation
Q1	What is the current status of Indian Machine translation systems	Identify the duration in which the large and important publications are done
Q2	Which approaches of machine translation are in use?	Identify the different approaches of machine translation development
Q3	What machine translation method has been used the most?	Identify the most popular and efficient technique for MTS development
Q4	What are the tools used for the method in Q3	Identify the most efficient tools and techniques used with their domains
Q5	What machine translation evaluation methods have been used the most?	Identify the most popular machine evaluation methods used largely and effectively
Q6	What new research avenues have obtained from the survey?	Explore new possible research avenues on which work needs to be done

Table 3 Search strategy

Sr. No.	Key phrase	Search string
1	History	Historical Background of MT
2	Approaches	Machine Translation Approaches
3	Corpus	Parallel, Aligned, Tagged Corpus
4	POS	Part of Speech Tagger
5	Statistical	Statistical Machine translation Systems
6	Rule Base	Rule Based MT Systems
7	Example Based	Example Based MT Systems
8	Direct	Direct Machine Translation Systems
9	UNL	Universal Networking Language Based MTS
10	NMT	Neural Machine translation Systems
11	ANN	Artificial Neural Network Based MTS
12	Parser	Different types of Language Parser
13	Evaluation of MT	Different methods of evaluating MTS
14	MT Challenges	Various challenges in MTS development
15	Semantic/syntactic analyzer	Natural language semantic/syntactic analyzers
16	Transfer rules	MT translation rule base

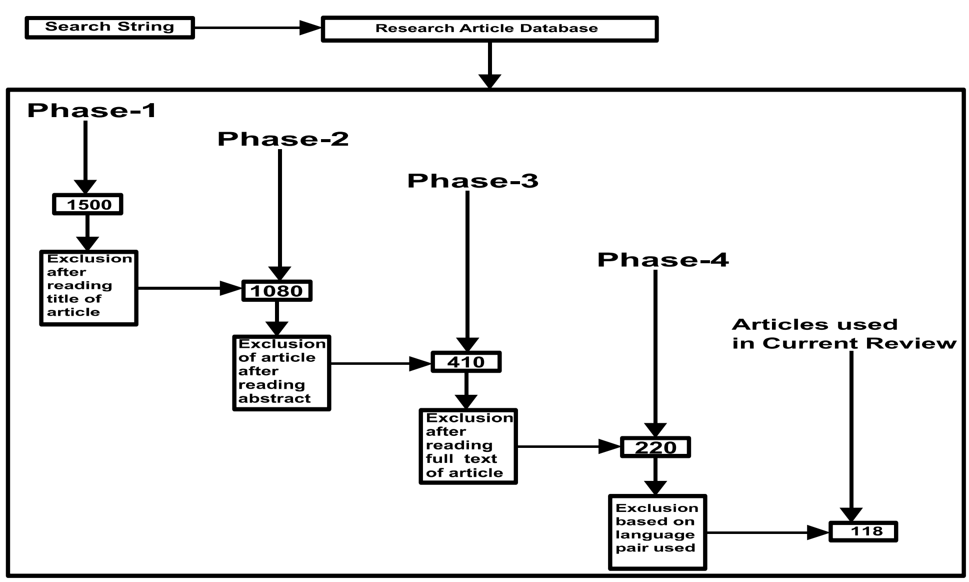

Fig. 11 Inclusion and or exclusion criteria

4 Results and discussion

This article examines the existing literature in the field of MT based on the research questions as per Table 2 and finds out the solutions to these questions as the outcome. Out of 118 articles, 45% are available in Journals, and 55% are published in conferences, workshops, Summits, Lecture Series and Technical Reports. The following subsections give an outcome-based analysis of various MTS and further examined based on approach, domain, and development year.

4.1 Machine translation system for Hindi and Sanskrit languages

Hindi and Sanskrit both belong to the Indo-Aryan language family which is a subgroup of the Indo-European language family. Both the languages are free word order and different from English which follows Subject–Verb–Object (SVO) word order. Hindi and Sanskrit both use the Devanagari script and shares many common features with each other.

Sanskrit is one of the oldest languages in the world and has been treated as a holy language in India. In the past, it was the language of educated people and used as a major language in communication, literature, education, administrative documents, and spiritual activities. The treasure of Sanskrit includes not only scientific, mathematical, philosophical, medical, poetry, and religious information but also India's spiritual as well as cultural aspects. Several languages have emerged from Sanskrit including Indian as well as foreign languages. The Sanskrit users have decreased gradually with time. Recently the Indian government and some non-governmental agencies have started to promote the Sanskrit language so that more people can be associated with this beautiful, spiritual, and most powerful language of the world. Several efforts have been made in developing Sanskrit language MTS all around the world. Based on Panini grammar several tools for Sanskrit language analysis, parsing, and generation tools have been developed by different research groups. Special Center for Sanskrit Studies at Jawaharlal Nehru University (Prof. Girish Nath Jha) New Delhi, University of Hyderabad (Dr. Amba Kulkarni), IIT Bombay (Prof. Pushpak Bhattacharya), IIT Kanpur (Prof. RMK Sinha and Pawan Goyal), Banaras Hindu University Banaras have been the core places for Sanskrit language processing tools development.

Hindi is regarded as the fourth most spoken language in the world and is also morphological rich (Lane 2016). Different research groups have been working to develop MTS for Hindi and Sanskrit languages following various MTS approaches. Tables 4 and 5 provide an overview of such MT systems based on several criteria which include approach used, year, language pair, features, domain, and efficiency. The next section discusses these systems based on the approach used for development and suggests solutions to improve their efficiency.

4.1.1 DMT based MTS

Based on the DMT approach three MTS have been included in this survey (Dubey 2019b; Dubey et al. 2013; Goyal and Lehal 2010). The main drawbacks of these MTS were that these systems were not able to resolve the word sense ambiguities, context resolution, translation of complex sentences because in the DMT approach word to word replacement strategy is followed. These issues can be resolved either by combining DMT with other approaches or by improving the lexicon of words with more syntactic as well as semantic attributes.

4.1.2 CBMT based MTS

Four MTS based on the CBMT approach have been included for review (Jain et al. 2001; Sachdeva et al. 2014; Sinha 2004; Sinha and Thakur 2005). The problems of NER, out of corpus translation in Jain et al. (2001) were resolved by Sinha (2004) adding special modules which will handle a particular problem. This modular approach makes the system more scalable and flexible. The problem of the polysemous verb with Sinha and Thakur (2005) can be resolved either by adding a special module as done in Sinha (2004) or by using the finite-state automaton approach or enhancing the POS tagger capability to resolve the issue. The issue with Sachdeva et al. (2014) is the feature extraction from the dataset which can be resolved easily with the help of deep neural networks (LSTM, RNN, CNN). Based on NMT citepmujadia-sharma-2020-nmt, kumar2019augmented, singh2020corpus, Laskar et al. (2020) systems have been developed. Evaluation of two MTS have also been covered (Goyal and Lehal 2009) and (Dungarwal et al. 2014). Other evaluation metrics like METEOR, NIST, R-L/W/S can be applied to validate these systems.

4.1.3 RBMT based MTS

Several MTS and MT tools have been considered for review based on the RBMT approach. The MTS using UNL as Interlingua were having issues of scalability and limited rule base which can be removed by the learning and feature extraction capabilities of neural networks even without the deep knowledge of SL and TL (Singh et al. 2007). The MTS based on GB theory was able to translate only simple

Table 4 Overview of Hindi MTS

MT approach	Year	MTS	Features	Efficiency	Domain	Citation
CBMT	1995	ANUBHARTI	Performs Hindi to English translation with abstracted example base to reduce size of corpus and simple distance based matching function for finding required output from the corpus Uses finite state machine for translation and can be used as base model to do translation among Indian languages	Better for translation among Indian languages	General	Antony (2013), Garje and Kharate, (2013), Jain et al. (2001)
	2004	ANUBHARTI-II	An extension of ANUBHARTI model which performs translation from Hindi-to Indian Languages and uses both Generalized Hierarchical Example Base over Rule Base with Automatic pre and post editing, Named Entity Recognition and Error Analysis Modules	Perform better than ANUBHARTI	General	Sinha (2004)
	2005	Hinglish	Uses one more layer over the ANUBHARTI-II and ANGLABHARTI-II MTS architecture and uses existing lexical databases, morphological analyzer, stemmers tools, Hindi verb endings for the MT development	> 90%	Narrations	Sinha and Thakur (2005)
	2009	Hindi-Punjabi MT Evaluation	Performs accuracy, intelligibility, Word Error Rate (WER) test on Hindi-Punjabi MTS. Uses daily news articles, literature, Blog data for testing	WER = 5.2% Intelligibility = 87.4%	Daily News, Blog and Literature	Goyal and Lehal (2009)

Table 4 (continued)

MT approach	Year	MTS	Features	Efficiency	Domain	Citation
	2014	Hindi–English MT	Uses phrase based and hierarchical based model for training. Uses GIZA + + and SRILM for phrase alignment and language model for training. Achieved minimum training error rate using MERT tool. Feature Vector and regression models are used to evaluate the quality of translation. Uses ILCI corpus for experiment purpose. Better than Google and Bing MTS	BLEU score = 21.82	Health	Sachdeva et al. (2014)
	2014	Hindi–English MT Evaluation	Uses WMT platform for evaluation of Phrase based Hindi–English and English–Hindi MTS. Uses Stanford Tokenizer and Moses tool for corpus normalization, shallow parser, small set of rule base to do translation for Hindi–English MTS and sentences of 50 word length has been used for testing purpose	Hindi–English BLEU score = 13.7 and for English–Hindi BLEU = 10.1	General	Dungarwal et al. (2014)
	2018	Hindi–English classification	Uses CNN for classification into three category	Satisfactory	Twitter, Facebook data	Mathur et al. (2018)
RBMT	2007	HinD	Generates Hindi text from UNL expressions using three simple steps. Uses UNL graphs to represent knowledge of natural language and UNL relations to remove ambiguities	BLEU Score = 0.34	Agriculture	Singh et al. (2007)
	2009	Hindi–English Translation	Uses Government and Binding Theory principles and Universal Grammar for translation. The phrase structure for Noun, Verb, Adjective, Preposition, Inflection and Complement phases have been developed. Hindi parse tree has been translated into English parse tree and node movement is done for word order problem resolution	Satisfactory	Simple sentences	Choudhary and Singh (2009)

Table 4 (continued)

MT approach	Year	MTS	Features	Efficiency	Domain	Citation
	2010	Sampark MTS	Hindi to Telugu, Tamil, Punjabi, Marathi, Bengali, Urdu, Kannada and Tamil to Telugu, Malayalam to Tamil Bidirectional MTS. Uses multi-column Shakti Standard Format for Input and Output purpose, Computational Paninian Grammar for handling free word order. Uses CRF++statistical tool for POS tagging	40% enhancement	General	Christopher and Rao (2010)
	2018	Hindi CCG Treebank	Uses two step process for extracting the sentences	96%	General	Ambati et al. (2018)
	2020	Hindi–Sanskrit MTS	Uses RBMT approach and common features of both languages	86%	General	Bhadwal et al. (2020)
	1994	Anusaaraka	Hindi-IL	Motivating	General	Narayana (1994)
DMT	2010	Web based Hindi–Punjabi MTS	Uses two digital data set have been built from Bhasha Vibhag and another from National Book Trust traditional dictionaries. Uses font converter to convert text of different fonts into Unicode and accept input from different sources like access, word, HTML, text file. Have 11 steps for translation with Web interface	95%	News articles	Goyal and Lehal (2010)
	2013	Hindi–Dogri MTS	Comparative analysis of Hindi and Dogri language results in similarity feature extraction in terms of script, grammar and word order, dissimilarity feature in terms of word inflections	98.5%	General	Dubey et al. (2013)
	2019	Hindi–Dogri MTS	Grammatical analysis of Hindi and Dogri language is performed to select approach of MT, pre-processing and handling of inflections in both SL and TL	98.71%	General	Dubey (2019a)
NMT	2020	Hindi–Marathi	Uses RNN seq2seq architecture for bidirectional translation among Hindi and Marathi language pair	BLEU score = 20.62%	General	Mujadia and Sharma (2020)

Table 4 (continued)

MT approach	Year	MTS	Features	Efficiency	Domain	Citation
HBMT	2011	Hindi–Punjabi MTS	Uses combination of DMT with RBMT approach for developing MTS and uses lookup, pattern matching algorithms to solve various difficulties like non-availability of the source language database, multiple spelling of the words in the source language, collocations faced while developing MTS. Uses Punjabi Unigram Wordnet to identify correct words	87.60%	Punjabi News	Goyal and Lehal (2011)
	2014	Hindi–English MTS	Uses hybridization of Quantum Neural Networks (QNN) with RBMT approach. Inputs are processed first with RBMT then with QNN architecture	BLEU Score = 0.7502	General	Narayan et al. (2014)
	2019	Hindi–English MTS	Uses IBMT and TBMT for translating Hindi Idioms to English	NA	General	Mishra et al. (2019)

sentences whose capability can be enhanced by the application of minimalist approach and generating the transfer rules either using SMT or NMT (Choudhary and Singh 2009). Hindi to Sanskrit and Sanskrit to Gujarati translation systems (Bhadwal et al. 2020; Raulji and Saini 2019) have been discussed. The efficiency of Sampark MTS was enhanced with the help of Memcached technique which can be done with LSTM network models (Christopher and Rao 2010). The Shakti Standard Format (SSF) format used in the system can be applied to other MTS which involves modular approach (Bharati and Kulkarni 2009). Two MTS for Sanskrit have also been included (Aparna 2005; Upadhyay et al. 2014). Several tools have been developed to process Sanskrit text (Bhadra et al. 2009; Kulkarni 2013; Kulkarni et al. 2010; Kumar et al. 2010). One issue regarding the morphological analysis of feminine nouns was reported by the authors to the developer in 2018 and that was rectified later on by the developer (Kulkarni 2013). The issues with these tools are that these are still in the testing phase. By developing the automatic testing tools for such systems an help in finding the issues early and fix them as soon as possible.

4.1.4 HBMT based MTS

Five MTS based on HBMT approach have been included for survey (Bawa et al. 2020a,b; Goyal and Lehal 2011; Narayan et al. 2014; Sitender and Bawa 2018). Different combinations of MT approaches DMT with RBMT, QNN with RBMT and RBMT with DMT have been used for the development of these systems, respectively.

4.1.5 MTS outcomes

After studying above mentioned Hindi and Sanskrit MTS thoroughly Figure 12 shows the possible outcomes.

4.2 Machine translation system for the English language to Indian languages

Several MTS have been proposed based on different approaches for English language which is the third most spoken language worldwide (Lane 2016). This section discusses such systems based on the approach used for development followed by a tabular representation of such systems is presented in Table 6.

4.2.1 RBMT based MTS

Based on RBMT approach, various MTS have been categorized into four groups. The first group have used pseduo-interlingua code (Goyal and Sinha 2009; Jayan and Bhadran 2014; Sinha and Jain 2003; Sinha et al. 1995; Sinha 2005) and second group has used UNL intermediate code

Table 5 Overview of Sanskrit MTS

MT approach	Year	MTS	Features	Efficiency	Domain	Citation
RBMT	2009	Sanskrit–Hindi	Performs translation from Sanskrit to Hindi language based on Anusaraka platform. Removes training problem for the user to understand the output. More user friendly. Takes input from different source file formats like pdf/html/text and produces out- put in Apertium format. Uses C Language Intergrated Production System for development of the system	Motivating	General	Bharati and Kulkarni (2009)
	2005	Sanskrit–English	Uses Morphological analysis, Sandhi rules and transducers for translation	Motivating	General	Aparna (2005)
	2009	Sanskrit Analysis System	Complete framework for analysing the Sanskrit Sentence. Di vided into Shallow Parser and Karaka Anlyser Module. Performs Sandhi, Samasa, Subanta, Gender, Kradanta, Taddhita, Tinanta	90%	General	Bhadra et al. (2009)
	2010	Sanskrit Com pound Processor	Describes the formation of compounds formation in Sanskrit language with four modules. Uses Sandhi and Optimality theory for segmentation and ranking in first module. For binding of segments it uses Constituency parser and type identifier for tagging. Uses Paraphrase generation for generating the compound	63%	General	Kumar et al. (2010)
	2010	Constraint based parser	It is a Sanskrit language parser. Uses four design principles for the parser development by following the grammar approach. Uses the graph representation of the input sentence. Uses 5D matrix representation for the implementation of the graphs	86%	Simple sentences	Kulkarni et al. (2010)
	2013	Deterministic parser	Uses dynamic programming concept for designing the parser for Sanskrit Language. Uses depth first search for resolving the relations among the nodes of the parse tree. Uses Sanskrit Tree Bank for the development	Relations with correct attachment UAS = 80.26%	Modern short stories	Kulkarni (2013)
	2014	TranSish MTS	Performs Sanskrit to English translation using RBMT approach. Uses simple transliteration of Sanskrit word with English word and apply reordering of words according to English grammar	Motivating	General	Upadhyay et al. (2014)
	2019	Sanskrit–Gujarati MTS	Uses RBMT approach for translating Sanskrit to Gujarati text. Uses constituent mapping of Sanskrit word with Gujarati word	BLEU score = 0.58	General	Raulji and Saini (2019)
SMT	2011	Sanskrit Corpus Tagger	Uses BIS tag set for tagging the corpus. Uses hierarchical structure for tagging process. Uses two main layers of the system with four noun category, six verb category and three conjunction cat	Motivating	General	Gopal and Jha (2011)

Table 5 (continued)

MT approach	Year	MTS	Features	Efficiency	Domain	Citation
	2016	SaHit	Analysis of errors in Sanskrit to Hindi translation. MTS is developed by using Microsoft Translation (MT) Hub. Uses 24k bilingual training set	BLEU Score = 41.17	Health, tourism	Pandey and Jha (2016)
NMT	2019	Sanskrit–Hindi MTS	Uses Zero Shot Translation method trained on English–Hindi and Sanskrit–Hindi data sets. Uses 300 Sanskrit-Hindi data set for testing the MTS	BLEU score = 13.3	News	Kumar et al. (2019)
	2020	Sanskrit–Hindi MTS	Uses CBMT approach with deep neural networks for translating Sanskrit to Hindi	BLEU score = 0.5	Bhagvad Geeta	Singh et al. (2020)
HBMT	2018	SANSUNL	Sanskrit to UNL translation using RBMT with DMT combination	BLEU Score = 0.85	General	Sitender and Bawa (2018)
	2020	Sanskrit to UNL enconverter	Uses LSTM for POS tagging and CFG for parsing	BLEU score = 0.81	General	Bawa et al. (2020b)
	2020	Sanskrit–English MTS	Uses hybridization of DMT and RBMT for translation	BLEU score = 0.7606	General	Bawa et al. (2020a)
GA	2019	Sanskrit–Hindi MTS	Uses Genetic Algorithm (GA) for translating text from Sanskrit to Hindi	Efficient than existing MTS	NA	Singh et al. (2019)s

to represent the intermediate code (Dave et al. 2001; Desai et al. 2014; Sridhar et al. 2016; Udupa and Faruquie 2005). The third group has translated the source syntax tree to target syntax tree using rule base (Aasha and Ganesh 2015; Bahadur et al. 2012; Darbari 1999; Pathak and Godse 2010). The fourth group uses Panini grammar rules, Sandhi rules, root word generation, pattern generation approach for translation (Ata et al. 2007; Balyan and Chatterjee 2015; Mishra and Mishra 2012; Reddy and Hanumanthappa 2013).

The issues with these systems are small size and non-standard form of analysis as well as generation rules, scalability, limited domain, time-consuming while writing the rules. The language processing tools like stemmer, POS tagger, parser used for the Indian language part were not competent with state-of-the-art tools like Porter stemmer, Malt parser, and Stanford parser. The approach followed in Porter stemmer to form the rule base should be adopted while making the rule base which will speed up the process. Language independent parsers should be developed like Malt parser or UNL parsers for Indian languages with the application of the NMT approach to remove the scalability and domain restriction issues.

4.2.2 CBMT and HBMT based MTS

Based on the CBMT approach several MTS have been proposed and classified into three groups. The first group has used statistical models like the IBM model, Bag of Words model, SRILM language model (OCH F 2007; Sharma 2011; Udupa and Faruquie 2005; Venkatapathy and Bangalore 2009). The second group has used Hierarchical phrase-based, simple phrase-based SMT techniques to perform the translation (Ali et al. 2013; Jawaid et al. 2014; Khan et al. 2013). The third group has used the EBMT approach for translation (Badodekar 2003).One system has also used the machine learning technique for the English–Bengali question–answer system (Sheikh and Conlon 2013). The issues with these are the availability of parallel aligned corpus of sentences, the complexity of statistical techniques to form the language as well as translation models which can be resolved with the help of the NMT approach or hybridization with other approaches. Application of machine learning techniques for prediction like CRF++, LSTM, RNN. Three MTS have been included based on the HBMT approach. Bharati et al. (2003) and NCST (2008) have used RBMT with SMT, while Narayan et al. (2014) have used RBMT with QNN for translation.

4.2.3 English MTS outcomes

Based on the discussion done in the above section and Table 6, Fig. 13 shows the outcomes obtained.

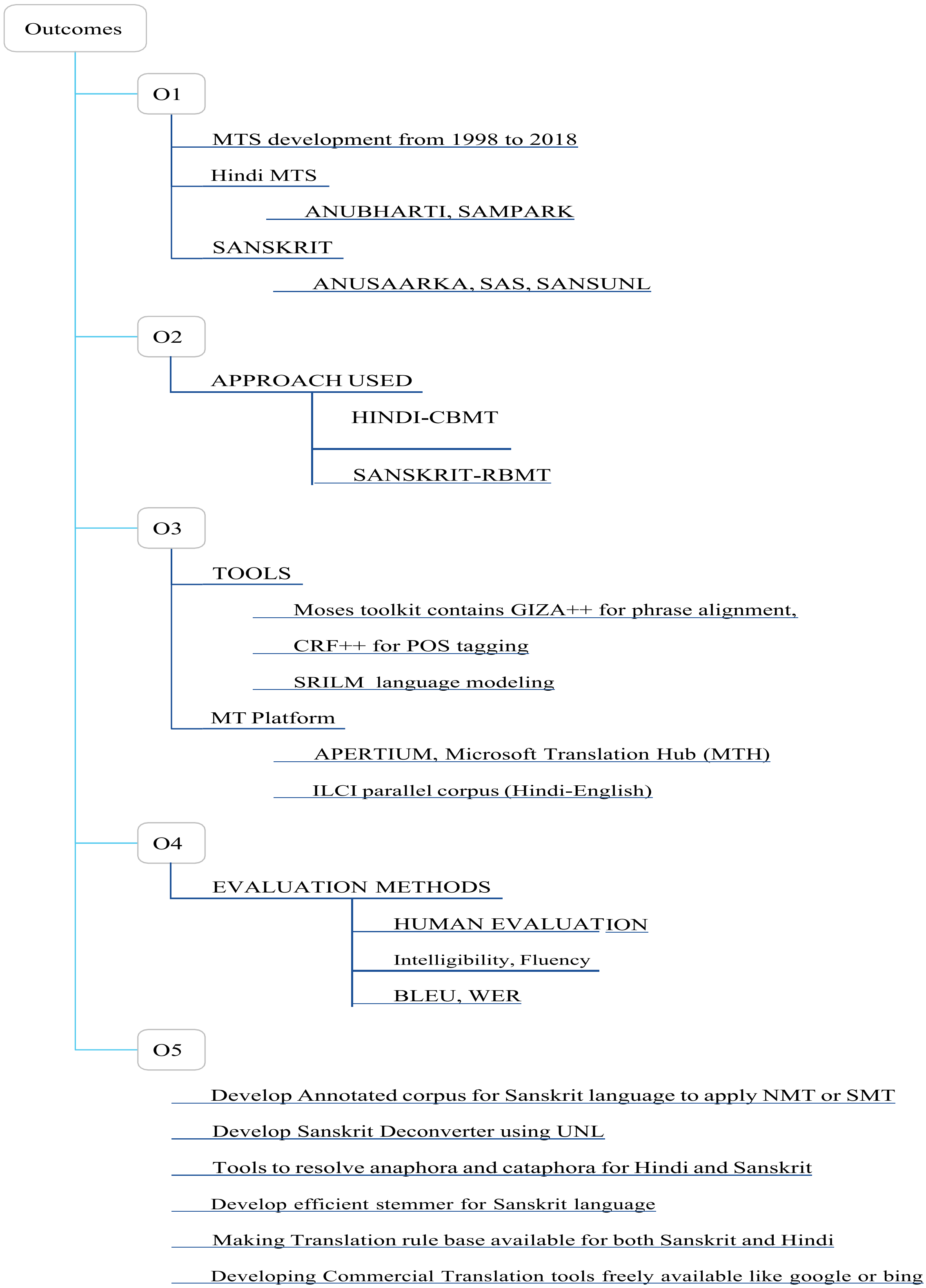

Fig. 12 Outcomes of Sanskrit and Hindi MTS

Table 6 Machine Translation System Based on English Language

MT Approach	Year	MTS	Features	Efficiency	Domain	Citation
RBMT	1995	ANGLABHARTI	Performs translation from English to Indian Languages. Uses CFG grammar and rule base for English language Analysis and generates pseudo-code as an intermediate code for translation. Uses human crafted rules for disambiguation	Motivating	General	Sinha et al. (1995)
	1999	MANTRA	Performs English to Hindi translation. Uses VYAKARTA parser for input text and KOSHAKAR tool for grammar generation Uses UNL as Interlingua for translation	95%	Admin Domain	Darbari (1999)
	2001	English-Hindi MTS	Uses UNL EnCo and DeCon, dictionary builder tools for development of MTS. Uses CYC On- tology with 3000 concepts to represent Language. Provides solution for language diver- gence problems	95%	Techno scientific	Dave et al. (2001)
HBMT	2003	ANGLAHINDI	Performs English to Hindi translation and extension of ANGLABHARTI system. Uses not only rule base but also example base and statistical approach for getting better performance. HMM could be used to rank the alternate translation	90%	General	Sinha and Jain (2003)
	2005	ANGLABHARTI- II	Enhancement of ANGLABHARTI system with addition of several Computer Assisted Tools (CAT) and CFG to convert the English text into PLIL. Uses translation memory, auto- matic pre and post editing tool, example base (raw, generalized), failure, paraphrasing mod- ules	40% enhancement	General	Sinha (2005)

Table 6 (continued)

MT Approach	Year	MTS	Features	Efficiency	Domain	Citation
RBMT	2007	English-Urdu	Translates text from English to Urdu language using CFG and Panini Grammar rules. Stanford parser is used for English language parsing. Recursive swapping of verb phrase in the parse tree is used to generate SOV format in Urdu	Motivating	General	Ata et al. (2007)
	2009	English-Malayalam	Uses RBMT approach with two rule bases one for morphological analysis and other for target language generation	limited to translate with 6 words sen tences	General	Rajan et al. (2009)
	2009	AnglaSanskrit	Performs English to Sanskrit Translation following the ANGLABHARTI project guidelines and Asthadhyayay rules. Able to handle affirmative, interrogative imperative, activeand passive voice sentences	Satisfactory	General	Goyal and Sinha (2009)
	2010	English-Sanskrit	Uses English parser for tree generation and using target language rules generate the equivalent parse tree for translation	Motivating	General	Pathak and Godse (2010)
	2012	Etrans	English to Sanskrit MTS Provides a complete framework for English to Sanskrit Translation. Uses Synchronous CFG to represent the language syntax. Uses both top down and bottom up approach for language translation model. 500 sentences were used for evaluation purpose	90%	General	Bahadur et al. (2012)
	2012	English-Sanskrit	Performs translation in eight steps. No parse tree is used for translation. Uses rule base engine to do the translation	BLEU Score = 0.551	General	Mishra and Mishra (2012)
	2013	English-Kannada/Telugu	Uses rule base and Dictionary base approach for translation	57%	General	Reddy and Hanuman-thappa (2013)

Table 6 (continued)

MT Approach	Year	MTS	Features	Efficiency	Domain	Citation
	2014	ANGLAMALAYALA	English to Malayalam Extension of Anglab-harti project. Uses pseudo-interlingua methodfor translation and also demonstrated the translation from English to Dravidian languages	75%	Health and Tourism	Jayan and Bhadran (2014)
	2014	English-Hindi MTS	Uses Stanford dependencies representation. Feature extraction is done using Morpha, RelEx and Function Tagger tools. Case Marking follows the feature transfer phase. CJ Re-ranking and Berkley parser are also discussed	Bleu score = 0.23, 0.18, 0.27 for different parser	Agriculture	Desai et al. (2014)
	2014	English–Hindi Noun Compound MTS	Uses semantic relations for translating noun compounds from English to Hindi. Uses 2-word semantic relation pattern recursively for translating 3-word or 4-word compounds. Uses 728 seed verbs and prepositions to identify semantic relation. Uses 20 semantic relations for the translation	83%	Literature	Balyan and Chatterjee (2015)
	2015	English–Malayalam	Uses Stanford parser for source language parsing and structure rules with bilingual dictionary for generating the target sentence	86%	Cricket match records	Aasha and Ganesh (2015)
	2016	English–Tamil MTS	Uses UNL inrterlingua for translatsing text from English to Tamil	BLEU Score=0.581	General	Sridhar et al. (2016)
SMT	2005	English–Hindi MTS	Uses IBM Model 1, 2, 3 for implementation. Uses 150,000 parallel sentence corpus for development. 1032 sentences are used for testing purpose	BLEU Score=0.1298	News, Government Documents	Udupa and Faruquie (2005)
	2007	Google Translator	Multi-lingual Translation using multi-engine among several languages. Currently showing translation among 90 languages	Motivating	General	OCH (2007)

Table 6 (continued)

MT Approach	Year	MTS	Features	Efficiency	Domain	Citation
	2009	English–Hindi MT	Uses global lexical selection for bi-directional translation purpose. Uses EILM tourism corpus for development and English sentences of word length 24, Hindi sentences with 26 words are used for testing purpose	BLEU Score E-H=0.1469 and for H-E=0.1483	Tourism information	Venkatapathy and Bangalore (2009)
	2011	English–Hindi SMT	Uses SRILM tool for language modeling and Giza++ tool with mkcls for translation model development. Uses Moses tool as decoder	Fluency=2.693 Adequacy=2.93	Freedom fighter history in court trail	Sharma (2011)
	2013	English–Urdu MTS	Uses Hierarchical Phrase Based model capability of strong generalization and reordering is used over EMILLE data base of parallel sentence for development and MERT tool is used for testing phase	BLEU Score=0.132	General	Khan et al. (2013)
	2013	Modeling English–Urdu MTS	Performs English to Urdu translation. Uses Sahih Bukhari and Sahih Muslim for dataset preparation. Provides solution for sentence alignment and discusses about phrase extraction problem while translation	BLEU Score=32.11	Ahadeeth translation	Ali et al. (2013)
	2013	DENGALI	Performs translation from English to Bengali. Uses Morphadorner POS tagger. Uses Point-wise Mutual Information (PMI) statistical technique for target language sentence selection	74%	Online travel guide	Sheikh and Conlon (2013)
	2014	English–Urdu SMT	Performs English–Urdu translation on Phrase Based as well as Hierarchical Modeling and tested on 3 test data. Uses Trrex platform for tokenization and lemmatization of source text. Uses 5-gram SRILM language model.	PBMT performs better than HMT	News wires and Web	Jawaid et al. (2014)

Table 6 (continued)

MT Approach	Year	MTS	Features	Efficiency	Domain	Citation
EBMT	2003	SHIVA	Translates English text into Hindi. Developed by Indian Institute of Science, Bangalore, India, Carnegie Mellon University USA and International Institute of Information Technology, Hyderabad	Motivating	General	Naskar and Bandyopadhyay (2005)
	2003	English–Hindi	Developed by IBM India Research Lab. Initially uses Example base and then uses SMT approach. Uses IBM language model and Giza++, Moses tool for statistical development	Language Model Score=21.27 and Translation score=8.36	General	Badodekar (2003)
	2008	English–Sanskrit MTS	Uses ENCG parser for English and Gerard Huet parser for Sanskrit language. Highlights the language divergence among English and Sanskrit language and also discusses the solution for the same	Motivating	General	Mishra and Mishra (2008)
	2012	Evaluation of English–Sanskrit	Proposes the weighs to BLEU score, Unigram precision, Unigram recall, F-measure and METEOR for evaluation. Weights are assigned to POS tags based on the reference translation and the proposed method translation	Improved evaluation score	General	Mishra and Mishra (2012)
NMT	2020	English–Hindi	Uses multi-model concept for translation	BLEU score=0.3357	WAT2020	Laskar et al. (2020)
HBMT	2003	SHAKTI MTS	Performs translation from English to Indian languages (Hindi, Marathi and Telugu). Uses RBMT and SMT approaches in hybrid form. Uses 69 modules divided into three phases	Motivating	General	Bharati et al. (2003)
	2005	EB-ANUBAAD	Performs English to Bangla language translation. Uses RBMT with TBMT for translation. Uses semantic nets for disambiguation	98%	General	Saha (2005)

Table 6 (continued)

MT Approach	Year	MTS	Features	Efficiency	Domain	Citation
	2008	MATRA	English to Hindi translation by using RBMT with SMT approach. 315 sentences are used for testing the system	BLEU Score = 0.0534	General	NCST (2008)
	2009	English–Sanskrit	Uses RBMT with ANN for translation. Uses feed forward Neural Networks. Uses Java and MATLAB for the implementation	BLEU Score = 0.445	General	Mishra and Mishra (2009)
	2011	English–Urdu MTS	Uses RBMT with feed forward neural network for translating text from English to Urdu. Uses Stanford parser and POS tagger for English language	BLEU Score = 0.6954	General	Shahnawaz and Mishra (2011)
	2014	English–Hindi MTS	Uses RBMT with Quantum Neural Network for translation	BLEU Score = 0.7502	General	Narayan et al. (2014)
	2015	English–Urdu	Uses Case Based Reasoning (CBT), RBMT and ANN for enhancing the translation efficiency. Uses Stanford type dependency parser for English language. Uses Levenberg–Marquardt back propagation algorithm for training the Feed forward Neural Network	BLEU Score = 0.728	General	Shahnawaz and Mishra (2015)
	2019	English–Urdu MTS	Uses Translation rules and ANN hybridization for translation	BLEU = 0.5903 METEOR = 0.7956	General	Khan and Usman (2019)

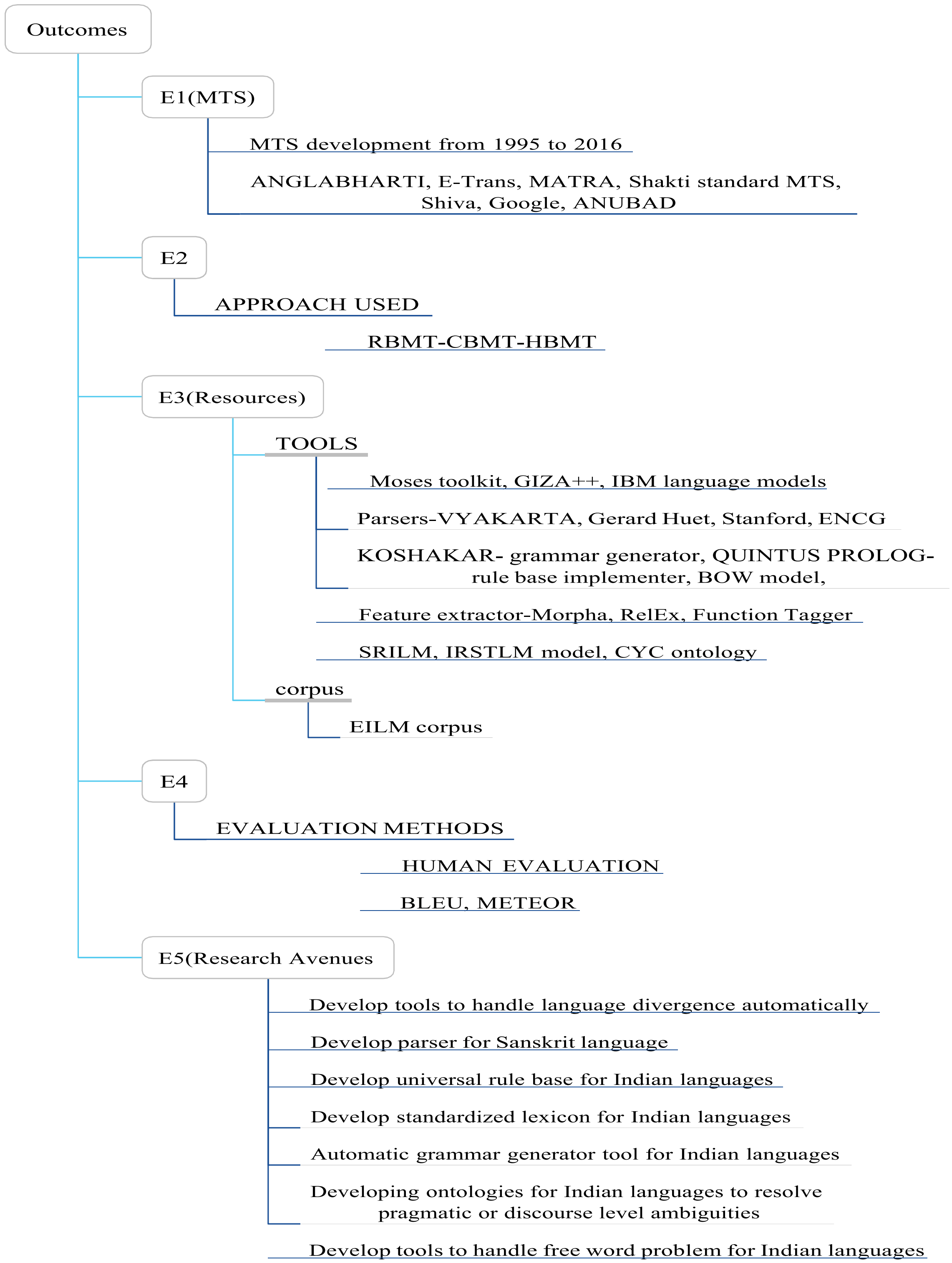

Fig. 13 Outcomes of English to Indian languages MTS

4.3 Research questions vs outcome

Ten outcomes are obtained after discussing the MTS in Subsects. 4.1 and 4.2 and are tabulated in Table 7. Research Questions are denoted by O1, O2, O3, O4, O5 and Q1, Q2, Q3, Q4, Q5, Q6 are the outcomes for Hindi and Sanskrit MTS while E1, E2, E3, E4, E5 are outcomes of English MTS. A four scale mapping is done with value ‘3’ as the

Table 7 Outcome and research questions

Outcome	RQ					
	Q1	Q2	Q3	Q4	Q5	Q6
O1	3	3	2	0	0	0
O2	1	2	3	0	0	0
O3	0	0	1	3	0	0
O4	0	0	0	0	4	0
O5	0	2	1	2	2	3
E1	3	3	0	0	0	0
E2	0	3	3	2	0	0
E3	0	0	0	3	3	0
E4	0	0	3	0	0	0
E5	0	2	2	2	2	3

maximum contribution and value of '0' indicates least contribution of an outcome with respect to the research questions as shown in Table 7.

5 Machine translation platforms and tools

This section gives an overview of some statistical tools, parser and corpus available online for developing new MTS and can be downloaded freely as shown in Table 8. Table 9 shows some of the popular MTS platforms which could be used for developing new MTS. Various language corpora available for Indian languages are also highlighted. Enabling Minority Language Engineering (EMILLE) contains three types of corpora such as parallel, monolingual and annotated. In parallel corpus it contains two lakhs words for Bengali, Gujarati, Hindi, Punjabi, and Urdu to English and reverses. Twenty annotated Hindi files are there in the corpus.

Gyan Nidhi corpus contains fifty thousand number of pages as a parallel corpus for each of eleven Indian languages including (Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Oriya, Punjabi, Telugu, Tamil) and English language.

Table 8 Online Resources

Resource	Citation
MTS	
Moses Statistical MTS	Koehn (2009)
Cunei Hybrid for Example Based and Statistical MTS	Phillips (2011)
Joshua Statistical MTS	Post et al. (2015)
Language Modeling Tool	
CMU-Cambridge Statistical Language Modeling Toolkit v2(Open Source)	Rosenfeld and Clarkson (1997)
SRILM ToolKit (Open Source) 7	Stolcke (2002)
IRSTLM Toolkit open source	Federico et al. (2008)
Neural Probabilistic Language Model Toolkit	Vaswani et al. (2013)
Neural Network Joint Model	Devlin et al. (2014)
Shallow Parser	
For Bengali, Hindi, Kannada, Malayalam, Marathi Punjabi, Tamil, Telugu	Hyderabad (2018)
Complete Parser	
Malt Parser (language Independent)	Nivre et al. (2007)
For Hindi, Tamil, Telugu, Urdu	Pune (2018)
Parallel Corpora	
EMILLE	Baker et al. (2002)
OPUS	Tiedemann (2009)
ILCI	Jha (2010)
Gyan Nidhi	Pune (2018)
Bilingual parallel sentences	Kelly (2021)

Table 9 Popular MTS Platform

MT platform	Language pair	Domain	Features	Organization	Citation
Google Translator	Multilingual	General	60% reduction in error of translation using GNMT	Google 2016	Wu et al. (2016)
Yandex Translator	Multilingual	General	More fluent and human like translation	Yandex	Yandex (2017)
Microsoft Translator Hub	Multilingual	General	Supports 60 language systems and 10 speech systems, produces netter results	Microsoft	Microsoft (2016)
OpenNMT	Language Independent Multilingual	General	Dependency free, simple, compatible to any language pair	Systran, Harvard nlp	Klein et al. (2017)
Stanford NMT	Multilingual	General	BLEU score of 5.2	Stanford University	Luong and Manning (2015)
Apertium Platform Open Source	Multilingual	General	Language Independent	Apertium	Forcada et al. (2011)

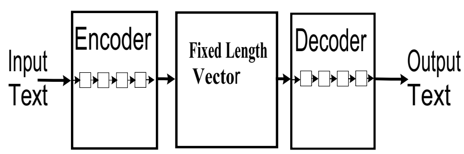

Fig. 14 NMT system architecture

Open Source Parallel Corpus (OPUS) contains parallel corpus for Assamese, Bengali, Bhojpuri, English, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Malayalam, Marathi, Oriya, Punjabi, Sanskrit, Tamil, Telugu and Urdu.

ILCI (Indian Language Corpora Initiative) contains a corpus of 50,000 parallel aligned sentences in Bangla, English, Hindi, Gujarati, Konkani, Malayalam, Marathi, Oriya, Punjabi, Urdu, Tamil, Telugu in the domain of tourism and health.

6 Role of artificial neural network in machine translation

With the explosive growth of the internet and easy access to high computing power systems, Neural Machine Translation has emerged as a fast-growing approach for developing new MTS (Cho et al. 2014; Kalchbrenner and Blunsom 2013; Sutskever et al. 2014).

The basic components of the NMT system are the encoder and decoder. It uses single neural network architecture to generate a target sentence for the input sentence, instead of using multiple small components optimized in pipeline form for obtaining translation in traditional phrase-based systems as shown in Fig. 14. Initially, the problem with NMT systems was the fixed- size vector space generated by the encoder for input sentence which was resolved by Bahdanau et al. (2014).

Different types of neural network architectures have been used for developing new MTS. Recurrent Neural Networks (RNN) are used mostly for MTS development due to their feature of preservation with the processing of input data/memorization of features of natural language. LSTM (Long Short-Term Memory) a type of RNN with two or more than two hidden layers is used for extracting features from the input text and increases the efficiency of translation (Agrawal 2017).

Machine Translation among eleven Indian languages using the NMT approach has been proposed and obtained better results than the traditional SMT approach (Agrawal 2017). Microsoft provided NMT based translation support for 21 languages and added Hindi recently (Microsoft 2017). Wu et al. (2016) also uses the NMT approach over the existing SMT approach and show better results than SMT. Facebook in 2017 proposed the implementation of NMT using Convolutional Neural Networks and claimed faster performance than the work presented by Gehring et al. (2016, 2017). Amazon has also launched its machine translation system using NMT approach (Faes 2018). Some important platforms useful for the development of NMT systems includes Tensorflow, Torch, Theano, PyTorch, Matlab, DyNet-lamtram and EUREKA are available at Zhang (2017).

7 MT evaluation methods

The MT evaluation methods are divided into two categories : Traditional Evaluation Methods and Automatic Evaluation Methods

Table 10 4 Point fluency score

Fluency score	4 point fluency score
1	Incomplete/not intelligible
2	Acceptable
3	Fair
4	Perfectly acceptable

Table 11 Sentence ranking by G Van Slype

Sentence	Rank
Sentences are unintelligibile	0
Sentences are having grammatical errors	1
Sentences are intelligible generally	2
Sentences are perfectly intelligible and clear	3

7.1 Traditional evaluation methods

This section will highlight some of the commonly used methods of MT evaluation (Van Slype 1979) following the traditional approach.

7.1.1 Fluency test

Fluency of an MTS gives the measure of the amount with which the target text is well-formed according to the TL grammar rules. A grammatically well-formed with correct spellings, stick to the common use of terms, names, and titles which can easily be interpreted and acceptable by the native speaker of the TL is known as the fluent segment (Singh et al. 2007; Goyal 2010). The 4-point scale was used in the evaluation of the Punjabi EnConverter and DeConverter System. The fluency score using Table 10.

7.1.2 Intelligibility evaluation

It provides the measure of easiness with which the translated text can be understood by the user. In this method, a group of persons is required to read the sentences in various versions (original, human translation with and without revision, MT without and with post-editing) in such a way that a particular person is receiving only one copy of the sentences of a particular version in the group. The ranking of the sentences on a 4-point scale is shown in Table 11 (Van Slype 1979). The ranking is received from the readers, and the average is taken of all the rankings to find out the overall intelligibility rank of the translation. This approach is applied to the evaluation of the Hindi–Dogri language, Hindi to Punjabi MTS, Punjabi to Hindi MTS, SYSTRAN English–French MT system. According to Carroll (1966) the measure of intelligibility is done on a 9-point scale as shown in Table 12.

This scale is used in the evaluation of automatic translation of ALPAC system.

7.1.3 Fidelity/adequacy test

Fidelity is the measure of an amount of information correctly translated into the TL from SL. It tells about the correctness of the translation. Rating of fidelity should be less than or equal to the intelligibility ratings and is done on a 4-point scale. It has been applied to the evaluation of Hindi–Dogri MTS, Punjabi Deconverter and English–French MT produced by the SYSTRAN system in which the rank of '3' means complete faithful and rank of '0' means completely unfaithful.

7.2 Automatic evaluation methods

Several automatic evaluation methods have also been proposed. Some of the popular methods are included for the survey and compared based on different metrics as shown in Table 13.

7.3 MT evaluation platforms

This section provides information about evaluation platforms available to evaluate MT systems on various metrics. Three platform ORANGE, Asiya, and IQMT have been explained in Table 14.

Table 12 Sentence Ranking by J Caroll

Sentence	Rank
Perfectly clear and intelligible sentence	9
Perfectly clear and intelligible sentence with minor grammatical mistakes	8
Generally clear and intelligible	7
The general idea is intelligible only after considerable study	6
Masquerades as an intelligible sentence, but actually it is more unintelligible than intelligible	4
Generally unintelligible	3
Almost hopelessly unintelligible	2
Hopelessly unintelligible	1

Table 13 Comparison of MT Evaluation Metrics

Criteria	BLEU	METEOR	NIST	WER	TER	ROUGH-L,W,S	RED	MaxSim
Meaning	Bilingual evaluation understudy	Metric for evaluation of translation with Explicit Ordering	National Institute of Standards and Technology	Word Error Rate	Translation Edit Rate	Longest Common Sub-sequence (LCS)	Reference Edit Distance	Maximum Similarity
Reference translated sentences required	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Word matching	Higher order n-grams	uni-gram matching	Weighted n-gram	Levenshtein distance	human annotations	in-sequence common n-grams	Human ranking encoding into vectors	Maximal weighted alignment matching framework
Mathematical Approach	Geometric mean of precision of n grams	Harmonic mean, recall and F-measure	Arithmetic mean of n gram counts	Dynamic programming to calculate WER	Dynamic programming to calculate number of edits	Minimum uni-gram F-measure	Decision Tree	Bipartite graphs for assigning different weights and match extraction
Final Score based on	Precision and penalty (numerical)	Precision, recall and F-measure	Precision and Penalty but weight assignment using different way	Editing distance among sentences	Quantitative Metric uses the ration	Precision and F-measure	Multiple distance instead of single	Uses n-grams and dependency relation matching
Complexity	less than others	More than BLEU	More than BLEU	Similar to BLEU	More expensive than others	less complex	more complex than ROUGH and WER	Equal to METEOR
Word form matching	Surface form only	Surface, stemmed form or meaning form of word	Surface and stemmed form	Surface form	Surface and stemmed	String matching	–	surface, stemmed and dependency relation level
Score Range	0 to 1	0 to 1	0 to 1	0 to 100%	0 to 1	0 to 1 for L, S and for W it is 1 to 9	A to Z	0 to 1
Better Translation score	Near to 1	Near to 1	Near to 1	Near to 0%	Near to 0	Near to 0	Near to a	Near to 1
Efficiency	Less at sentence level	Better than BLEU	Better than BLEU	–	Better than BLEU and METEOR	Better than BLEU	Better than BLEU	Better than BLEU, NIST and BLEU
Sentence Length	Neither too long nor short	No barrier	Sentence segments should be lesser	–	–	No barrier	No barrier	No barrier

Table 14 MT evaluation platforms

ORANGE (Lin and Och 2004)	It is an Oracle ranking for Gisting Evaluation. It does not require any human involvement other than the reference translation. It is used to evaluate different MT metrics in a better way. It requires only a single parameter optimization than other systems. Smaller the value of ORANGE the better the metric
Asiya (Giménez and Márquez 2010)	It is an open toolkit that allows the mixing of different metrics to estimate the quality of MT as well as the metric useful for a particular MT. It generates reports of MT evaluation based on four schemes (Model, QUEEN, single, and UIC). It is developed using Perl. Meta Evaluations use five different criteria (Spearman, Pearson, King, Kendall, and ORANGE). Several metrics like WER, TER, BLEU, ROUGE, METEOR, and NIST
IQMT (Gimenez and Amig 2006)	It is based on QARLA framework and is available at http://www.lsi.upc.edu/~nlp/IQMT. It uses three schemes for the evaluation report as Jack, Queen, and King. Several MT metrics like PER, WER, NIST, BLEU, GTM, and ROUGE have been used for evaluation

8 Research avenues and recommendations

Although lots of work have been done in the last three decades for developing MTS with different language pairs (Indian languages) and of various domains. The emergence of the NMT approach and the easy availability of high computing resources and corpus for Indian languages has created several new opportunities for researchers to work in this field. The researchers are now more focused to apply the machine learning algorithms for text processing rather than other fields and as a result, several new tools and platforms are available for text processing. It is a very difficult and time-consuming process to create the rule base which will cover all the aspects of the language specifically for Hindi and Sanskrit languages which are highly inflected and morphological rich in nature. To apply the SMT approach the need for a large corpus is again a big hurdle for languages like Sanskrit. The following are some of the research avenues with which the researchers can start their research work:

- Developing POS tagger or stemmer for Hindi and Sanskrit languages using a hybrid approach of rule base and machine learning techniques.
- Developing automatic Karaka Analyzer (case marker) for Sanskrit and Hindi by making use of the similarity features among Indian languages in such a way that only a small effort is required to make this system for other Indian languages.
- Developing a platform like Snowball (http://snowball.tartarus.org) for creating the rule base in an easy and fast manner.
- Creating small modules which can enhance the performance or reduce the response time of the existing MTS like the Named Entity Recognition (NER) tool, automatic pre- or post-processing tools using machine learning techniques.
- Anaphora or Catphora resolution is still a challenging task for the Sanskrit language. So, special modules can be developed for such types of problems which can be easily merged with the MTS adopting modular approach.
- For MTS using UNL as an interlingua approach, the resolution of UNL relation is a challenging area because it requires thousands of rules to resolve all the 56 UNL relations (Le Thuyen and Hung 2016). So, machine learning approaches can be used over the UNL dictionary to predict the possible relations with the Case marker module.
- Development of the Sanskrit Deconverter using UNL is still an open area of research.
- Development of Operating Systems for computers using less ambiguous language like Sanskrit.
- Developing tools to extract text from scanned images and develop digital corpus for languages like Sanskrit and Punjabi.

Based on the discussions done in Sects. 4.1, and 4.2 and the outcomes shown in Figs. 12, 13 on various MTS the following recommendations are derived for researchers working in field of machine translation:

- The application of any architecture (approach) to develop new MTS depends on various parameters like language pair, availability of linguistic resources for the language pair, the application domain of MTS, linguistic knowledge.
- SMT approach performs better for long sentence translation and DMT gives better results for short length sentences.
- Maximum utilization of similarity feature at syntax level or semantic level among Indian languages such as noun, verb, declension, prefix, Karka Analysis for case identification, word formation, and word order, etc. should be done for developing MTS among Indian Languages.
- Interlingua approach needs fewer efforts for developing multilingual MT systems like Anglabharti, Anubharti, UNL based MTS, and Sampark. So, Interlingua representation like of pseudo-Interlingua, UNL expressions, or an intermediate representation of Sanskrit language as Interlingua could be used efficiently for developing

new MTS, and less effort is required for new language translator development.
- Panini Grammar is one of the most unambiguous grammars ever developed for a natural language and written in a more structured manner for Indian languages. Panini principles will help to develop new MTS for Indian Languages based on the RBMT or HBMT approach.
- RBMT systems require deep linguistic knowledge of the source as well as the target language and are a time-consuming process although the quality of translation using RBMT is better than other approaches.
- Use of statistical tools like Moses' toolkit, Giza + +, IRSTLM, SRILM makes the developing process much faster than other systems but requires a large amount of parallel corpus in digital format, so applicable only for language pairs having large corpus availability in digital form.
- Google and Microsoft have used deep neural networks over the SMT approach and proved that the Neural Machine Translation approach performs much better than SMT and even requires fewer amounts of data for training, but requires large computational power to train such systems.
- For Sanskrit Language, various part of speech taggers is available like BIS POS, JPOS (JNU), CPOS, IL POS (Indian Language), and Gerard Huet Parser, Constraint-Based Parser, Deterministic Parser of Amba Kulkarni, and Indic NLP Library could be used to develop Sanskrit Based MTS.
- For English Language Stanford Parser is efficient enough to give the analysis of the English Language.
- The availability of wordnet for English, Hindi and Punjabi and Punjabi makes the translation task easier and less time- consuming. The shallow parser available on the TDIL website could be used for Indian Languages.
- The fastest way of developing MTS is by using the DMT approach, and the quality of translation is also good but limited to a small domain and requires bilingual dictionaries and a small number of transfer rules like in Sampark MTS.

The Hindi and Sanskrit languages have used the traditional methods of MT evaluation which include Fluency Test, Intelligibility Test, and Fidelity Test. Most of these tests depend on human evaluation but the application of the NMT approach be easily applied to them also. In the case of automatic evaluation methods, the BLEU and METEOR score has become the common standards for MT evaluation. For English to Indian language MTS the BLEU, NIST, and METEOR have been used by the developers.

9 Conclusion

This article presents an outcome-based systematic survey of machine translation for English, Hindi, and Sanskrit languages. Out of 1500 research articles, 118 articles have been included in this survey based on the Inclusion-Exclusion criteria mentioned in Subsect. 3.3. The results of the survey are presented in different dimensions like MT Evolution, MT approaches, mapping research questions with outcomes, overview of MTS based on several criteria (approach, language pair, domain, efficiency, features), state-of-the art-MT tool-kits, technological enhancement in MT approach, MT evaluation methods and platforms. The latest trends in MTS development are based on neural networks and provides human-like translation quality as seen in Hassan et al. (2018). Also, it is still not feasible for languages like Sanskrit to develop an efficient MTS and apply SMT or NMT approach due to non-availability of corpus and complexity of the language. State-of-the-art MTS platforms with MT development tools and corpus have also been discussed. State-of-the-art MT evaluation methods and platforms with specific features have been explored in this survey. Several research avenues have been highlighted in this survey work for further research in machine translation. Future recommendations have also been included to help researchers to develop new MT or enhance existing MT development.

Declarations

Conflict of interest We have no conflicts of interest to disclose.

Human and animal rights This article does not contain any studies with animals performed by any of the authors. This article does not contain any studies with human participants or animals performed by any of the authors.

References

Aasha V, Ganesh A (2015) Machine translation from English to Malayalam using transfer approach. In: Proceedings of international conference on advances in computing, communications and informatics (ICACCI), pp 1565–1570

Agrawal R (2017) Towards efficient neural machine translation for indian languages. PhD thesis, International Institute of Inforsmation Technology, Hyderabad

Ali A, Hussain A, Malik MK (2013) Model for English-Urdu statistical machine translation. World Appl Sci 24:1362–1367

Allen J (1995) Natural language understanding, 2nd edn. Pearson, London

Ambati BR, Deoskar T, Steedman M (2018) Hindi ccgbank: A ccg treebank from the Hindi dependency treebank. Lang Resour Eval 52(1):67–100

Antony P (2013) Machine translation approaches and survey for indian languages. Int J Comput Linguist Chin Lang Process 18(1):47–78

Aparna S (2005) Sanskrit to English translator. Lang India 5:1

Ata N, Jawaid B, Kamaran A (2007) Rule based English to Urdu machine translation. In: Proceedings of conference on language and technology, pp 1–7

Badodekar S (2003) Translation resources, services and tools for indian languages. Computer Science and Engineering Department, Indisan Institute of Technology, Mumbai

Bahadur P, Jain A, Chauhan D (2012) Etrans—a complete framework for English to Sanskrit machine translation. In: Proceedings of international conference and workshop on emerging trends in technology in international journal of advanced computer science and applications (IJACSA), Citeseer, pp 52–59

Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:14090473

Baker P, Hardie A, McEnery T, Cunningham H, Gaizauskas RJ (2002) Emille, a 67-million word corpus of Indic languages: data collection, markup and harmonisation. In: Proceedings of the international conference on language resources and evaluation (LREC), pp 819–825

Balyan R, Chatterjee N (2015) Translating noun compounds using semantic relations. Comput Speech Lang 32(1):91–108

Bawa S et al (2020a) A sanskrit-to-english machine translation using hybridization of direct and rule-based approach. Neural Comput Appl 33:2819–2838

Bawa S et al (2020b) Sanskrit to universal networking language enconverter system based on deep learning and context-free grammar. Multimedia Syst 1–17

Bhadra M, Singh SK, Kumar S, Agrawal M, Chandrasekhar R, Mishra SK, Jha GN et al (2009) Sanskrit analysis system (sas). In: In Kulkarni A., Huet G. (eds) Sanskrit computational linguistics. ISCLS 2009. Lecture notes in computer science. Springer, pp 116–133

Bhadwal N, Agrawal P, Madaan V (2020) A machine translation system from Hindi to Sanskrit language using rule based ap- proach. Scalable Comput Practice Experience 21(3):543–554

Bharati A, Kulkarni A (2009) Anusaaraka: An accessor cum machine translator. Department of Sanskrit Studies, University of Hyderabad, Hyderabad, pp 1–75

Bharati RM, Reddy P, Sankar B, Sharma D, Sangal R (2003) Machine translation: the Shakti approach. In: Proceedings of international conference on natural language processing (ICON-2003)

Budgen D, Brereton P (2006) Performing systematic literature reviews in software engineering. In: Proceedings of the 28th international conference on software engineering, ACM, pp 1051–1052

Carroll JB (1966) An experiment in evaluating the quality of translations. Mech Transl Comp Linguistics 9(3–4):55–66

Cho K, Van Merriënboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder-decoder approaches. arXiv preprint arXiv:14091259

Choudhary A, Singh M (2009) Gb theory based Hindi to English translation system. In: Proceedings of 2nd IEEE international conference on computer science and information technology (ICCSIT-2009). IEEE, pp 293–297

Christopher M, Rao UM (2010) IL-ilmt sampark: a hybrid machine translation system. In: Proceedings of 32nd all India conference of linguistics (AICL32). Lucknow University, Lucknow, pp 69–75

Darbari H (1999) Computer-assisted translation system—an Indian perspective. Machine Translation Summit VII, 13–17 September, pp 80–85

Dave S, Parikh J, Bhattacharyya P (2001) Interlingua-based English–Hindi machine translation and language divergence. Mach Transl 16(4):251–304

Desai NP, Dabhi VK (2021) Taxonomic survey of Hindi language nlp systems. arXiv preprint arXiv:210200214

Desai P, Sangodkar A, Damani OP (2014) A domain-restricted, rule based, English-Hindi machine translation system based on dependency parsing. In: Proceedings of the 11th international conference on natural language processing, pp 177–185

Devlin J, Zbib R, Huang Z, Lamar T, Schwartz R, Makhoul J (2014) Fast and robust neural network joint models for statistical machine translation. In: Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 1: long papers), vol 1, pp 1370–1380

Dorr BJ, Hovy EH, Levin LS (2004) Natural language processing and machine translation encyclopedia of language and linguistics, (ell2). Machine translation: interlingual methods. In: Proceeding of international conference of the world congress on engineering, pp 1–20

Dubey P (2019a) The Hindi to Dogri machine translation system: grammatical perspective. Int J Inf Technol 11(1):171–182

Dubey P (2019b) The Hindi to Dogri machine translation system: grammatical perspective. Int J Inf Technol 11:1–12

Dubey P et al. (2013) Machine translation system for Hindi-Dogri language pair. In: Proceedings of international conference on machine intelligence and research advancement (ICMIRA-2013), IEEE, pp 422–425

Dungarwal P, Chatterjee R, Mishra A, Kunchukuttan A, Shah R, Bhattacharyya P (2014) The IIT Bombay Hindi-English translation system at wmt 2014. In: Proceedings of the ninth workshop on statistical machine translation, association for computational linguistics, pp 90–96

Echizen-Ya H, Araki K, Momouchi Y, Tochinai K (2004) Machine translation using recursive chain-link-type learning based on translation examples. Syst Comput Jpn 35(2):1–15

Faes F (2018) Amazon and lion bridge share stage to market neural machine translation. https://slator.com/technology/amazon-and-lionbridge-share-stage-to-market-neural-machine-translation/

Federico M, Bertoldi N, Cettolo M (2008) Irstlm: an open source toolkit for handling large scale language models. In: Proceedings of ninth annual conference of the international speech communication association, pp 1618–1621. https://github.com/irstlm-team/irstlm

Forcada ML, Ginestı-Rosell M, Nordfalk J, O'Regan J, Ortiz-Rojas S, Perez-Ortiz JA, Sanchez-Martınez F, Ramırez-Sanchez G, Tyers FM (2011) Apertium: a free/open-source platform for rule-based machine translation. Mach Transl 25(2):127–144

Fromkin V, Rodman R, Hyams V (2011) An introduction to language, 9e. Wadsworth, Cengage Learning, Boston, MA

Garje G, Kharate G (2013) Survey of machine translation systems in India. Int J Nat Lang Comput (IJNLC) 2(4):47–67

Gehring J, Auli M, Grangier D, Dauphin YN (2016) A convolutional encoder model for neural machine translation. arXiv preprint arXiv:161102344

Gehring J, Auli M, Grangier D, Yarats D, Dauphin YN (2017) Convolutional sequence to sequence learning. arXiv preprint arXiv:170503122, pp 1–15

Gimenez J, Amig E (2006) Iqmt: a framework for automatic machine translation evaluation. In: Proceedings of the international conference on language resources and evaluation (LREC), pp 685–690. http://www.lrec-conf.org/proceedings/lrec2006/

Giménez J, Márquez L (2010) Asiya: an open toolkit for automatic machine translation (meta-) evaluation. Prague Bull Math Linguist 94:77

Gopal M, Jha GN (2011) Tagging Sanskrit corpus using bis pos tagset. In: Information systems for Indian languages, Springer, pp 191–194

Goyal V (2010) Development of a Hindi to Punjabi machine translation system

Goyal V, Lehal GS (2009) Evaluation of Hindi to Punjabi machine translation system. arXiv preprintarXiv:09101868

Goyal V, Lehal GS (2010) Web based hindi to punjabi machine translation system. J Emerg Technol Web Intellig 2(2):148–151

Goyal V, Lehal GS (2011) Hindi to Punjabi machine translation system. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies: systems demonstrations. Association for computational linguistics, pp 1–6

Goyal P, Sinha RMK (2009) A study towards design of an English to Sanskrit machine translation system. In: Sanskrit computational linguistics, Springer, pp 287–305

Hassan H, Aue A, Chen C, Chowdhary V, Clark J, Federmann C, Huang X, Junczys-Dowmunt M, Lewis W, Li M, et al. (2018) Achieving human parity on automatic Chinese to English news translation. arXiv preprint arXiv:180305567

Hutchins WJ (1995) Machine translation: a brief history. Concise history of the language sciences: from the Sumerians to the cognitivists, pp 431–445

Hutchins WJ, Somers HL (1992) An introduction to machine translation, vol 362. Academic Press, London

Hyderabad I (2018) Machine translation and natural language processing lab. http://ltrc.iiit.ac.in/

Jain R, Sinha R, Jain A (2001) Anubharti-using hybrid example-based approach for machine translation. In: Proceedings of symposium on translation support systems (STRANS-2001), IIT, Kanpur, pp 20–32

Jawaid B, Kamran A, Bojar O (2014) English to Urdu statistical machine translation: establishing a baseline. In: Proceedings of the fifth workshop on South and Southeast Asian natural language processing, pp 37–42

Jayan V, Bhadran V (2014) Anglabharati to anglamalayalam: an experience with English to Indian language machine translation. In: Proceedings of international conference on contemporary computing and informatics (IC3I), pp 282–287

Jha GN (2010) The tdil program and the Indian language corpora initiative (ilci). In: Chair NCC, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (eds) Proceedings of the international conference on language resources and evaluation (LREC), European Language Resources Association (ELRA), Valletta, Malta, pp 982–985. http:// sanskrit.jnu.ac.in/ilci/index.jsp

Jung H, Yuh S, Kim T, Park S (1999) A pattern-based approach using compound unit recognition and its hybridization with rule-based translation. Comput Intell 15(2):114–127

Kalchbrenner N, Blunsom P (2013) Recurrent continuous translation models. In: Proceedings of the 2013 conference on empirical methods in natural language processing, pp 1700–1709

Kelly C (2021) Tab-delimited bilingual sentence pairs from the tatoeba project (good for anki and similar flashcard applications). https://www.manythings.org/anki/

Khan S, Usman I (2019) A model for English to Urdu and Hindi machine translation system using translation rules and artificial neural network. Int Arab J Inf Technol 16(1):125–131

Khan N, Waqas A, Bajwa U, Durrani N (2013) English to urdu hierarchical phrase-based statistical machine translation. In: Proceedings of international joint conference on natural language processing, Japan, pp 72–76

Kitchenham B, Brereton OP, Budgen D, Turner M, Bailey J, Linkman S (2009) Systematic literature reviews in software engineering—a systematic literature review. Inf Softw Technol 51(1):7–15

Klein G, Kim Y, Deng Y, Senellart J, Rush AM (2017) Opennmt: open-source toolkit for neural machine translation. arXiv preprint arXiv:170102810

Koehn P (2009) Moses–statistical machine translation system

Kulkarni A (2013) A deterministic dependency parser with dynamic programming for Sanskrit. In: Proceedings of the second international conference on dependency linguistics (DepLing 2013), pp 157–166

Kulkarni A, Pokar S, Shukl D (2010) Designing a constraint based parser for Sanskrit. In: Sanskrit computational linguistics. Springer, pp 70–90

Kumar A, Mittal V, Kulkarni A (2010) Sanskrit compound processor. In: Sanskrit computational linguistics, Springer, pp 57–69

Kumar R, Jha P, Sahula V (2019) An augmented translation technique for low resource language pair: Sanskrit to Hindi translation. In: Proceedings of the 2019 2nd international conference on algorithms, computing and artificial intelligence, pp 377–383

Lane J (2016) The 10 most spoken languages in the world. URL https://www.babbel.com/en/magazine/the-10-most-spoken-languages-in-the-world/

Laskar SR, Khilji AFUR, Pakray P, Bandyopadhyay S (2020) Multimodal neural machine translation for English to Hindi. In: Proceedings of the 7th workshop on Asian translation, pp 109–113

Le Thuyen PT, Hung VT (2016) Automatic translation for vietnamese based on unl language. In: 2016 international conference on electronics, information, and communications (ICEIC), IEEE, pp 1–5

Lewis MP, Simons GF, Fennig CD (2015) Ethnologue: languages of Ecuador. SIL International, Texas

Lin CY, Och FJ (2004) Orange: a method for evaluating automatic evaluation metrics for machine translation. In: Proceedings of the 20th international conference on computational linguistics, association for computational linguistics, pp 501–507

Luong MT, Manning CD (2015) Stanford neural machine translation systems for spoken language domains. In: Proceedings of the international workshop on spoken language translation, pp 76–79

Mallikarjun B (2010) Patterns of Indian multilingualism. Strength Today Bright Hope Tomorrow 10(6):1–18

Mathur P, Shah R, Sawhney R, Mahata D (2018) Detecting offensive tweets in Hindi-English code-switched language. In: Proceedings of the sixth international workshop on natural language processing for social media, pp 18–26

Microsoft (2016) Microsoft translator launching neural network based translations for all its speech languages. https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/

Microsoft (2017) Microsoft translator accelerates use of neural networks across its offerings. https://blogs.msdn.microsoft.com/translation/2017/11/15/microsoft-translator-accelerates-use-of-neural-networks-across-its

Mishra V, Mishra R (2008) Study of example based english to sanskrit machine translation. J Res Dev Comp Sci Eng 37:1–12

Mishra V, Mishra R (2009) Ann and rule based model for English to Sanskrit machine translation. INFOCOMP J Comput Sci 9(1):80–89

Mishra V, Mishra R (2012) English to sanskrit machine translation system: a rule-based approach. Int J Adv Intellig Paradig 4(2):168–184

Mishra H, Chakrawarti RK, Bansal P (2019) Implementation of Hindi to English idiom translation system. In: International conference on advanced computing networking and informatics, Springer, pp 371–380

Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA (2015) Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement. Syst Rev 4(1):1–9

Mujadia V, Sharma DM (2020) Nmt based similar language translation for Hindi-Marathi. In: Proceedings of the fifth conference on machine translation, pp 414–417
Narayana V (1994) Anusarak: a device to overcome the language barrier. PhD thesis, Ph. D. thesis, Dept. of CSE, IIT Kanpur
Narayan R, Singh V, Chakraverty S (2014) Quantum neural network based machine translator for Hindi to English. Sci World J 2014:1–8
Naskar S, Bandyopadhyay S (2005) Use of machine translation in india: Current status. AAMT J 36:25–31
NCST (2008) Matra: an English to Hindi machine translation system. Tech. rep., NCST MUMBAI
Nivre J, Hall J, Nilsson J, Chanev A, Eryigit G, Kübler S, Marinov S, Marsi E (2007) Maltparser: a language-independent system for data-driven dependency parsing. Nat Lang Eng 13(2):95–135
OCHF (2007) Google translator. In: Proceedings of joint conference on empirical methods in natural language processing and computational natural language learning, association for computational linguistics, Prague, pp 858–867
Pandey RK, Jha GN (2016) Error analysis of sahit-a statistical Sanskrit-Hindi translator. Proc Comp Sci 96:495–501
Pathak G, Godse S (2010) English to Sanskrit machine translation using transfer approach. In: Proceedings of international conference on methods and models in science and technology. American Institute of Physics, Pune, pp 122–126
Phillips AB (2011) Cunei: open-source machine translation with relevance-based models of each translation instance. Mach Transl 25(2):161–177
Post M, Cao Y, Kumar G (2015) Joshua 6: a phrase-based and hierarchical statistical machine translation system. Prague Bull Math Linguist 104(1):5–16
Pune C (2018) Indian language technology proliferation and development centre. http://tdil-dc.in/index.php?lang=en
Rajan R, Sivan R, Ravindran R, Soman K (2009) Rule based machine translation from English to Malayalam. In: Proceedings of international conference on advances in computing, control, & telecommunication technologies, 2009. ACT'09, IEEE, pp 439–441
Rao DD (1998) Machine translation. Resonance 3(7):61–70
Raulji JK, Saini JR (2019) Sanskrit-Gujarati constituency mapper for machine translation system. In: 2019 IEEE Bombay section signature conference (IBSSC), IEEE, pp 1–8
Reddy MV, Hanumanthappa M (2013) Indic language machine translation tool: English to Kannada/Telugu. In: Multimedia processing, communication and computing applications, Springer, pp 35–49
Rosenfeld R, Clarkson P (1997) Cmu-cambridge statistical language modeling toolkit v2
Sachdeva K, Srivastava R, Jain S, Sharma DM (2014) Hindi to English machine translation: using effective selection in multi-model smt. In: Proceedings of the international conference on language resources and evaluation (LREC), pp 1807–1811
Saha GK (2005) The ebanubad translator: a hybrid scheme. J Zhejiang Univ Sci A 6(10):1047–1050
Seasly J (2003) Machine translation: a survey of approaches. University of Michigan, Ann Arbor
Shahnawaz A, Mishra R (2011) Translation rules and ann based model for English to Urdu machine translation. INFOCOMP J Comput Sci 10(3):25–35
Shahnawaz A, Mishra R (2015) An English to Urdu translation model based on cbr, ann and translation rules. Int J Adv Intellig Paradigms 7(1):1–23
Sharma N (2011) English to Hindi statistical machine translation system. PhD thesis, M. Tech. thesis, Thapar University Patiala
Sheikh M, Conlon S (2013) Application of machine translation in bilingual knowledge management. Int J Intercult Inf Manage 3(2):123–137
Singh S, Dalal M, Vachani V, Bhattacharyya P, Damani OP (2007) Hindi generation from interlingua. In: Proceedings of machine translation summit, pp 1–8
Singh M, Kumar R, Chana I (2019) Ga-based machine translation system for Sanskrit to Hindi language. In: Recent trends in communication, computing, and electronics, Springer, pp 419–427
Singh M, Kumar R, Chana I (2020) Corpus based machine translation system with deep neural network for Sanskrit to Hindi translation. Proc Comp Sci 167:2534–2544
Sinha R, Jain A (2003) Anglahindi: an English to Hindi machine-aided translation system. In: Proceedings of MT Summit IX, New Orleans, USA, pp 494–497
Sinha RMK (2004) An engineering perspective of machine translation: anglabharti-ii and anubharti-ii architectures. In: Proceedings of international symposium on machine translation, NLP and translation support system (iSTRANS-2004), pp 10–17
Sinha RMK (2005) Integrating cat and mt in anglabharti-ii architecture. In: Proceedings of the 10th European association for machine translation (EAMT) conference, pp 235–244
Sinha RMK, Thakur A (2005) Machine translation of bi-lingual Hindi-English (hinglish) text. In: Proceedings of the 10th machine translation summit (MT Summit X), Phuket, Thailand, pp 149–156
Sinha R, Ivaraman K, Agrawal A, Jain R, Srivastava R, Jain A et al (1995) Anglabharti: a multilingual machine aided translation project on translation from English to Indian languages. In: Proceedings of IEEE international conference on systems, man and cybernetics. Intelligent systems for the 21st century, IEEE, vol 2, pp 1609–1614
Sitender, Bawa S (2018) Sansunl: a Sanskrit to unl enconverter system. IETE J Res 1–12. https://doi.org/10.1080/03772063.2018.1528187
Slocum J (1985) A survey of machine translation: its history, current status, and future prospects. Comput Linguist 11(1):1–17
Sridhar R, Sethuraman P, Krishnakumar K (2016) English to Tamil machine translation system using universal networking language. Sādhanā 41(6):607–620
Stolcke A (2002) Srilm—an extensible language modeling toolkit. In: Proceedings of seventh international conference on spoken language processing, pp 1–4. http://www.speech.sri.com/projects/srilm/
Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Proceedings of advances in neural information processing systems, pp 3104–3112
Tiedemann J (2009) News from opus—a collection of multilingual parallel corpora with tools and interfaces. In: Recent advances in natural language processing, vol 5, pp 237–248. http://opus.nlpl.eu/
Udupa R, Faruquie TA (2005) An English-Hindi statistical machine translation system. In: Natural language processing–IJCNLP 2004, Springer, pp 254–262
Upadhyay P, Jaiswal UC, Ashish K (2014) Transish: translator from Sanskrit to English—a rule based machine translation. Int J Curr Eng Technol E-ISSN, pp 2277–4106
Van Slype G (1979) Critical study of methods for evaluating the quality of machine translation. Prepared for the Commission of European Communities directorate general scientific and technical information and information management report BR 19142
Vaswani A, Zhao Y, Fossum V, Chiang D (2013) Decoding with large-scale neural language models improves translation. In: Proceedings of the 2013 conference on empirical methods in natural language processing, pp 1387–1392
Venkatapathy S, Bangalore S (2009) Discriminative machine translation using global lexical selection. ACM Trans Asian Lang Inf Process (TALIP) 8(2):8

Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K et al (2016) Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint arXiv:160908144

Yandex (2017) Yandex blog. https://yandex.com/company/blog/one-model-is-better-than-two-yu-yandex-translate-launches-a-hybrid-machine-translation-system/

Zhang M (2017) History and frontier of the neural machine translation

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.